

CLASSROOM TESTS



RUSSELL

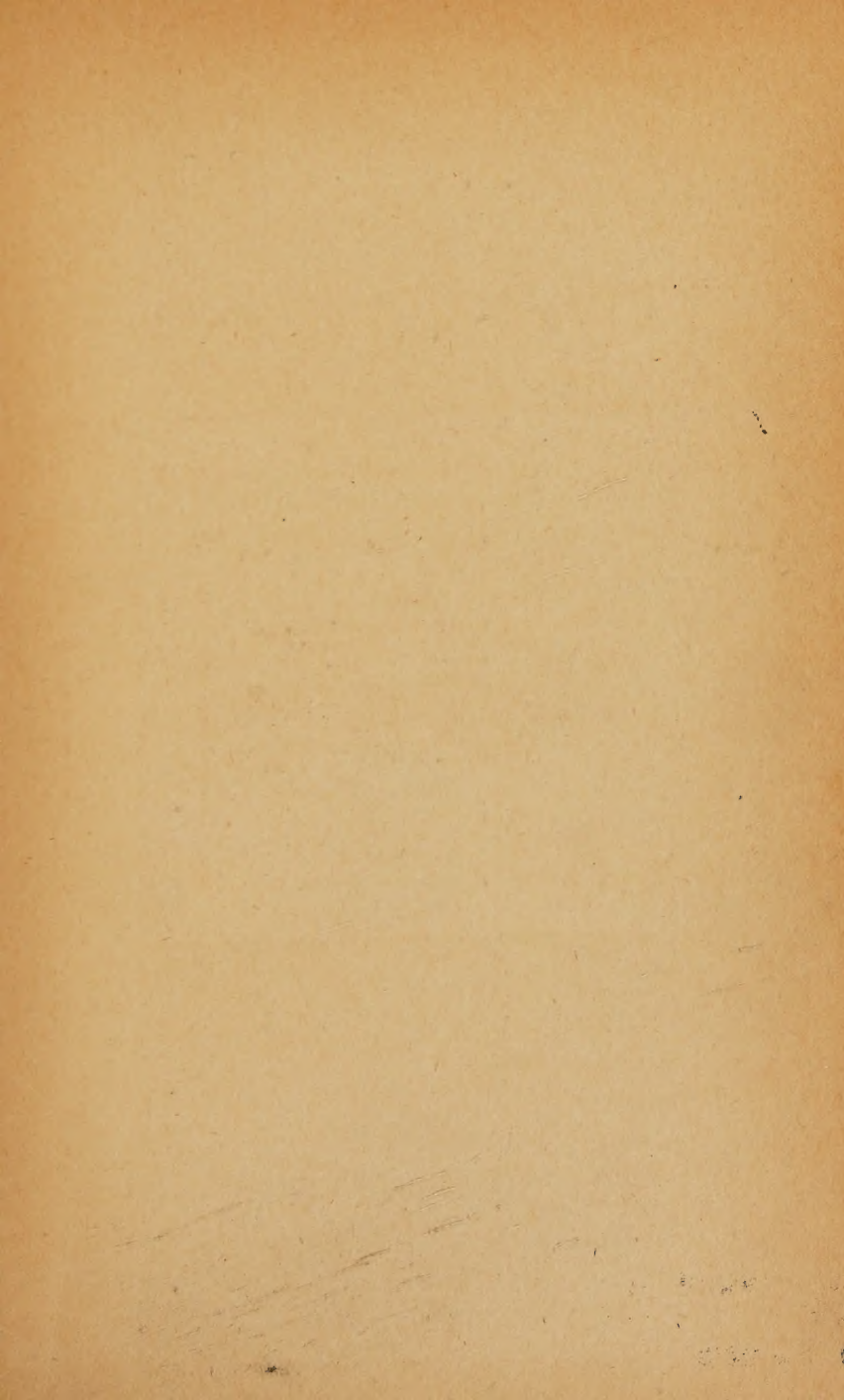
LIBRARY

OHIO NORTHERN UNIVERSITY

REGULATIONS

1. Not more than two books **can** be taken out at any one time and cannot be retained longer than two weeks.
2. At the expiration of two weeks books may be re-let if returned to the librarian in good condition.
3. A rental of two cents a day will be charged on each book held over two weeks.

Class No. 377.26 Accession No. 11140



CLASSROOM TESTS

A HANDBOOK ON THE CONSTRUCTION AND
USES OF NON-STANDARD TESTS FOR
THE CLASSROOM TEACHER

BY

CHARLES RUSSELL, PH. D.

PRINCIPAL, THE MASSACHUSETTS STATE NORMAL
SCHOOL AT WESTFIELD



GINN AND COMPANY

BOSTON · NEW YORK · CHICAGO · LONDON
ATLANTA · DALLAS · COLUMBUS · SAN FRANCISCO

COPYRIGHT, 1926, BY CHARLES RUSSELL

ALL RIGHTS RESERVED

PRINTED IN THE UNITED STATES OF AMERICA

826.2

The Athenæum Press

GINN AND COMPANY • PROPRIETORS • BOSTON • U.S.A.

PREFACE

A group of teachers from a single school in Toledo, who had attended a late-afternoon university class conducted by the writer and who had been subjected to the type of testing given in this book, began to adapt the methods for use in their own classrooms. Encouraged by their principal and aided somewhat by the writer, they developed early forms of the tests here described. The immediate results of the work were generally good, and in a few instances of outstanding interest. For example, in one upper-grade classroom, two boys, uninterested in the school program, difficult to handle, bent upon completing school at the earliest possible moment, and unmoved by the efforts of the teacher, were suddenly galvanized into sustained school effort by the challenge of the tests. Crude as were these early tests, they proved themselves none the less a promising teaching-device.

The use of the tests spread to other teachers and other schools, and the methods of making and interpreting them became more and more refined. Many types of test were tried, some of them good, others of little value, and many processes and procedures were developed. The tests and methods that are given in the following pages are those which have proved successful. They have been the response to definite needs of the teachers who have used the tests, and have been developed from specific classroom situations. They may be criticized, perhaps, from many points of view, but their great value lies in the fact that they are and have been useful, that they have added something to teaching.

Because such a book is not, and cannot be, the work of a single individual, the writer wishes to express his appreciation to those who have coöperated in the various phases of

its production. Among these should be mentioned Superintendent Charles S. Meek and Assistant Superintendent Estaline Wilson, both of the Toledo public schools, and former President A. M. Stowe, of the University of Toledo, who are responsible for the necessary administrative adjustments which have made the work possible; Dr. William A. McCall, who gave the writer his first knowledge of testing and whose influence can be seen throughout the chapters which follow; Dr. William C. Bagley, whose encouragement and help have been unremitting; Dr. Thomas D. Wood, whose contributions have added much to parts of the book; Dr. William F. Russell, who has read the manuscript and has improved it greatly through his kindly criticisms; and several unknown readers, whose marginal notes have been of unfailing help. Finally, to Miss Jessica Marshall, principal of Newton School, Toledo, to the teachers of her school, and to those teachers in his classes who have coöperated so whole-heartedly with him and without whom the work would have been impossible, the writer wishes to express his appreciation and gratitude.

CHARLES RUSSELL


CONTENTS

PART ONE. WHY AND HOW TO MAKE TEACHER'S CLASSROOM TESTS

CHAPTER	PAGE
I. THE USES OF TESTS AND TESTING	3
II. THE TYPES OF TESTS USED IN SCHOOLS	12
III. THE TRUE-FALSE TEST	28
IV. THE JUDGMENT TEST	66
V. THE SELECTION TEST	89
VI. THE ASSOCIATION TEST	124
VII. THE COMPLETION TEST	147
VIII. MAKING TRADITIONAL-TEST TYPES EFFECTIVE . .	166
IX. THE USE OF BATTERIES OF TESTS	178

PART TWO. WHY AND HOW TO USE TEACHER'S CLASSROOM TESTS

X. THE DISTRIBUTION OF TEST SCORES	201
XI. THE MEANING OF CURVES	213
XII. THE DETERMINATION OF QUESTION DIFFICULTY .	226
XIII. THE USE OF TESTS FOR EDUCATIONAL DIAGNOSIS AND THE IMPROVEMENT OF TEACHING	244
XIV. THE MAKING OF COMPOSITE TEST SCORES	267
XV. JUDGING PUPILS IN ACHIEVEMENT ACCORDING TO ABILITY	283
XVI. JUDGING AND CLASSIFYING PUPILS ACCORDING TO GROUP PLACEMENT	301
XVII. THE USE OF CLASSROOM TESTS AS DEVICES IN TEACHING	329
INDEX	341



Digitized by the Internet Archive
in 2022 with funding from
Kahle/Austin Foundation

CLASSROOM TESTS

PART ONE. WHY AND HOW TO MAKE TEACHER'S CLASSROOM TESTS

FOREWORD

The construction of a building has much more in it than the erection of the ironwork, the laying of the brick, and the finishing of the interior. Before the steel can be ordered, or the brick brought, or the interior even considered, there is much to be done. First of all the site for the building must be selected, and then the architects must plan a building to fit. It should be no larger than the site; if smaller, then in harmony with it. The building that is planned should not be merely walls and roof; it should have character, and that character in terms of the purpose for which it is to be used. If it is a factory, that is one thing; if it is a hotel or a bank, that is another; if it is a school, it is still another. And then, before the building is even begun, the plans for its completion in terms of its purposes must be carefully drawn, and the necessary materials to meet those plans must be procured. Then, and only then, can construction proceed so that in the end the finished building will be most fully useful.

Likewise, before useful tests can be made there is much that the teacher can and should do. Like the buildings of industry, these buildings of education have their purposes, and whatever is planned should be planned to fit. The following chapters are designed to show how these tests may be made, and examples are given to show certain types of completed units, from which teachers may be enabled to derive some help and inspiration. As in many other types of construction, however, the greatest satisfaction will follow if every teacher is his own architect.

CHAPTER I

THE USES OF TESTS AND TESTING

Measurement in education. Among the more significant recent tendencies in education has been that of measuring the results of teaching. Theories in education have ever kept in advance of practice, particularly in elementary education, largely because the educational philosopher could name desirable results which the educational practitioner could not measure and therefore could not show had been attained. The past few years, however, have seen a remarkable development in the science of measurement as applied to education, — a development which has had two outstanding characteristics. One of these has been the increasing refinement and complexity of the measures which have been used; the other has been the ever-increasing uses to which these refined measures have been put.

The earlier and cruder measures in education were restricted, because of the inaccuracies which were the inevitable result, not only in their interpretation but in their application as well. The examinations of a former day served a good purpose, but they led to misinterpretations and misuse. More recent measures of the results of teaching show far more clearly the paths which a teacher may follow with surety and the means which he can use with increased efficiency.

These newer measures, together with the more traditional ones, may conveniently be called *tests*; and the uses to which they may all be put are merely extensions and refinements of former uses, made possible because of greater accuracy in the testing, greater pointedness in the results, and greater refinements in the measuring units.

Tests may measure memory. Tests may be used to measure the memory of pupils. There may be several types of memory tested. The type of test most frequently used in school is that which determines the amount of material an individual has retained. John has been studying about Africa. The test is one so designed as to determine how much John still remembers of what he was supposed to have learned about Africa. This is a good test of memory of pure knowledge; but, with the emphasis which is now being placed upon other types of memory and upon other attributes of knowledge, the test is, in this respect, somewhat too narrow. It is frequently good to remember things not as separate facts but as related facts. In this case some other type of test is needed: one which not only measures a knowledge of the facts themselves but also an additional knowledge of their relation to one another. It may also be desirable on occasion merely to have a memory which recognizes a fact, without necessarily having the ability to recall the fact unassisted. To recognize the truth of a fact, with the fact out of the context in which it may be usually found, constitutes a universal need of daily life and may on occasion prove as valuable as any other form of memory. The same may be said of the order of a series of facts. It is often as necessary for an individual to know the order of a series of facts as to have a memory for the facts themselves. All these types of memory should be used in school, and we should encourage them, in addition merely to stimulating the remembrance of facts supposedly "memorized" or "learned."

Tests may measure a teacher's efficiency. Tests may also be used as a measure of a teacher's efficiency. It is true we cannot say that a teacher has any definite percentage of teaching efficiency, as we can measure the development of horse power in a gasoline engine, and it would probably be undesirable if we could. We can, however, through the analysis of a series of tests, say that a teacher is doing better with his pupils than teachers of similar children elsewhere, or

as well, or more poorly. This, then, may become for the individual teacher a potent factor in helping him to appreciate where improvements in his teaching can be made and where effort should be best placed. Without necessarily knowing the absolute amount of efficiency with which a teacher teaches, it becomes possible in this way to increase efficiency by preventing unnecessary repetitions of teaching or by providing more clearly for teaching-needs.

Tests may be used for examinations. One of the widest uses which all forms of testing have had in the past, and one of their more important uses in the future, lies in the field of examinations. In this sense tests are used to find out the status of an individual pupil at any given time from the point of view of his achievement. It may be that the primary purpose of examining is to determine the fitness of a pupil for promotion. As long as promotion to a higher grade depends largely on the academic fitness of the individual for such promotion, — that is, as long as pupils are graded largely on the sum total of what they have acquired in the way of knowledge, — just so long is this method of examining for promotion just and right. The great need, however, is to make sure that the examinations which are used for this purpose really measure what is wanted and do measure real achievement. Examinations to determine the status of pupils who have just entered a school are also much needed. This, again, is a form of determination of the status of a pupil, not primarily as a check after teaching has been done, but rather to find out where further teaching should begin. In these forms of examinations, tests can be used to advantage in securing not only useful results but results which are justified.

Tests may be used for review and recall. One of the wider uses of testing is that of testing for purposes of review or to help in the process of recall. Psychologists have made many studies of the rate at which various subjects and skills are forgotten after they have once been learned. In conformance

with the Thorndike Laws of Learning it has been found that without exception whatever is learned tends to deteriorate, other things being equal, with disuse, but that also each review or recall makes the learned bonds stronger and stronger, until, with enough repetition of the right kind, the bonds may become so well established above the threshold of memory as to be relatively permanent. Under such conditions it would seem to be one of the great aims of the teacher not only to teach, but also to reteach so as to increase the retention of the learned elements. As will be shown later, some types of teacher's tests help to make the subject matter so vivid and so full of interest that it may not only be retained but also be retained in a psychologically desirable way. In addition to retention, however, there are other purposes in a review which are important in teaching and which can be aided through testing. One of these consists of organization and reorganization. It seems true that not only should things be learned in one way, but they frequently should be learned in others as well if they are to be really useful. Learning things in other ways means merely making different applications of them or of finding new relationships to them. This is reorganization. Testing can be used to great advantage for achieving this purpose.

Tests help in placement and classification. Another of the uses of testing which is being rapidly developed at the present time is for the proper placement of new pupils in a school and the reclassification of pupils already there. On the one hand, pupils coming into a new school district or into a new school system may have been taught by different standards, or possibly with a different curriculum, from the pupils in the schools which they are entering. Tests can be used to advantage with these prospective pupils to determine not only their absolute achievement but also their abilities in relation to the other pupils in the school. The result of these tests provides justification for making whatever placements are made. On the other hand, many pupils, under the former

bases of promotion, were frequently misplaced several grades, although they were more often placed in grades below their actual school level than above it. In this case a battery of tests administered throughout a school may be an objective determiner of the status of all pupils in the school and it may also be a valid reason for changing the grade classification of many of them.

Diagnosis is a worthy use of tests. In addition to these uses, however, one of the more worthy uses to which testing can be put lies in diagnosis. This is a field as broad as that of the school itself, but one to which until the present there has been but little attention paid. Diagnosis of the difficulties in the learning process, of the difficulties for individual pupils in the varying types of subject matter and in the various phases of the school curriculum, all this is one of the widest and most important fields of endeavor for the classroom teacher. It is a field of constant inquiry, a field of highly focused endeavor, and a field of rapid change. Here it is that the teacher has to do with the kaleidoscopic changes of tastes and perception, of likes and dislikes, of feelings and emotions. Here it is that he has to work with the individual differences of his pupils and with their inherent abilities. No adequate method by which these differences could be detected, no usable means by which these changes could be measured, no fair treatment of the pupils of a class, was possible, until tests and the method of measurement as we know it today were developed. The diagnosis of individual difficulties and the remedial measures which must be taken to counteract these difficulties lie within the province of the classroom teacher. They can be accomplished only through the medium of adequate testing.

Tests may be used for comparison. One of the results of the testing which is now being done in the elementary schools is that a teacher may compare his class and the work of its individual members with like pupils in like classes in other places. In a nation as large as this — a nation where the

ideals of education are (in their more fundamental aspects at least) so universal in all sections, and where the character of the sections and the character of the people differ to such an extent — it is of great value that a teacher should know what may be the standards in other parts of the country and be able to compare his standards and his pupils with these. Tests may furnish the only reliable means by which this can be done.

Tests enhance the intrinsic worth of learning. Another use of tests is one which has yet had but little attention, but which is rich in possibilities for both the present and the future. It helps to realize one of the great ideals of elementary education; namely, to enhance for this and for future generations the intrinsic worth of learning. At a time when the efforts of educators are focused upon an ideal of worth-while activity on the part of pupils in school, when the curriculum is being scanned to remove traces of artificiality, when courses of study are being planned to eliminate subjects or parts of subjects that are included largely because of prejudiced tradition, and at a time when the work of the pupil in school is directed toward making his life there more rich and more like life outside of school, any plan which will enhance the intrinsic worth of learning to the pupil is a step toward the realization of those ideals. Some learning has passed from the stage of coercion through the stage of reward, past the stage of rivalry, and is now founded on the worth of learning for its own sake. Tests rightly used, emphasizing the worth of learning, the desirableness of knowledge, especially in terms of its usability, are means to bring more of the materials of education to this level. We can see the reasons for the ineffectiveness of rewards and sugar-coating as a basis for right learning. We can see the unwholesome ideals connected with emphasizing the rivalry of one's fellows. It should be clear that it is in rivalry of one's best previous efforts, of oneself, that real education results. In daily living rivalry of one's fellows seems, from a superficial

point of view, to be the paramount motive for success. But one needs only to examine the conspicuous cases of success in his own environment to discover that these are not merely the result of a selfish rivalry of the success of others. The conspicuously successful physician, lawyer, pastor, or teacher has no rival save himself. If this is the criterion for success in life, it should be the criterion for success in school. Tests rightly used and rightly interpreted furnish a means by which pupils can rival their own best previous efforts. It is a means of promoting one of the highest types of social education.

Use of tests to give pupils objective standards. A further use of tests is one which tends to give to pupils some objective standard by which to judge the character or the quality of the work which they have done. Pupils immersed in the details of school study frequently find it difficult to appreciate the objectives of that study. For that reason they have little, save their own interest or inclination, by which to judge the relative importance of varying phases of a subject as they arise. Tests may furnish such objectives (or at least a high grade of substitute for such objectives) by helping students to see relative values and to appreciate the necessity for accurate and complete knowledge. In another way the same object may be accomplished through the fact that the tests may provide a motive for the study. If the motive is not that of passing the subject, which is so common in the traditional type of test (though even there it need not be), it is probable that it is worthy and capable of furnishing a worthy objective. This objective, even if it is in some cases not the best which the educational philosopher would advocate, is, nevertheless, better than has been achieved in many cases in the past. At all events, tests may furnish objective standards and objective goals that are superior to non-objective standards or unknown goals.

Use of tests to improve teaching. Tests may be used in several ways to improve teaching. One way may be con-

sidered in its relation to the pupil. Tests furnish for the teacher a more complete knowledge of his pupils. Through testing he can discover the individual status of the pupils, together with their individual difficulties and misunderstandings. He can also discover the difficulties and misunderstandings that are characteristic of like pupils. When the teacher anticipates these difficulties in his teaching, he is improving it. A second way may be in the subject matter itself. Tests may be used to increase the value of materials already in use by giving to them wider application or greater implication, and they may also be used as guides to needed or desirable extensions of these materials. Either of these uses of tests should result in improved teaching. A third way of improving teaching may be in method. Tests furnish objective results of the methods by which teaching is accomplished. These objective results are not alone a measure of the pupil and his accomplishment; they are also a measure of the effectiveness of the method which the teacher has used. When varying methods are contrasted in terms of the results secured, and the better methods are chosen for future use, better teaching is one result. A fourth way in which teaching may be improved is through the teacher himself. Here the improvement may result from the wider knowledge which the testing makes necessary, or from the greater skill and confidence in the teaching which the tests make possible, or, most important of all, from the added stimulation to constructive thinking which accompanies the testing.

Chapter summary. Tests are being used in many ways besides merely for purposes of examination and promotion. They may be used to test the efficiency of a teacher or to examine pupils for purposes of locating beginning points in teaching or for determining school status. They may promote review and recall either by increasing the retentiveness of pupils or by organizing and reorganizing the work that has been covered. They may be used for placement of pupils or for classification. They may be given in order to

diagnose difficulties of pupils, and they may in that way provide a basis for remedial measures. They may be used so as to compare a class or a pupil in one part of the nation with another in a different part, or with the composite pupil of all parts. They may be used for the motivation of school work, for the promotion of real and not artificial interest, bringing with that the handmaiden of interest — active attention. They may provide pupils with objective goals for school study. Most important of all, to the extent to which their influence may be felt, tests may improve teaching.

CHAPTER II

THE TYPES OF TESTS USED IN SCHOOLS

The traditional school test. Of the tests now employed, whose possible results have been described in the previous chapter, that form which has been in use longest, the predecessor of all later forms, may be called the traditional school test. This appears in many different ways in educational institutions and is used on every level of instruction. In the classroom it appears in the form of the question, from which have descended the recitations of the question-and-answer type, teaching-methods that involve the question, and numerous problems of educators. At the other end of the oral scale, in its highest form, this test appears in all solemnity as the oral examination for the doctorate. In written form this test is found on the one hand in the well-known written examination used in schools, and at the other end of the scale in that formal test of research, experiment, or development, the doctor's dissertation.

Those who have passed through our elementary schools know that the written school tests have been used extensively and largely for the purpose of determining the worthiness of individuals for promotion, or for advancement within a grade. The tests are usually made by the teacher of the classroom; though in some places, notably in New York State, this examination has been constructed by state authorities for universal state use. The test has variable characteristics, though the most usual form consists of a series of questions based upon a certain unit or series of units of subject matter. The questions are usually of two types. One of these is the essay form, using such directions as "Describe," "Tell about," "Discuss," or "Criticize"; the

other consists of fact-questions preceded by "How" or "What" or "Why" and sometimes by "Where" or "Who."

Advantages of the traditional school test. This test has several great advantages. It is easy to construct. All a teacher has to do is to examine the subject matter that was taught and make a series of questions relating to it. The character of these questions, their degree of difficulty, the extent of knowledge which their answering requires, and the extent of the subject matter which they cover are all matters for the judgment of the teacher to determine. The test is also easy to give. Written on the blackboard or dictated by the teacher, it is made available for all the pupils, who may write their answers on paper or, as is sometimes done, in "examination books." Such an examination allows a great range of individual choice as to how the questions may be answered, and at the same time allows many different interpretations and levels of answers. On the higher levels of examining, as in colleges and universities, this is a distinct advantage, whereas on the lower levels of instruction it is frequently questionable. Perhaps the greatest advantage of this type of test is that it is directly adapted to the subject matter which the pupils have been taught. It may not cover this subject matter adequately, it may merely touch on certain obscure points; but it is based on this matter and, from that standpoint, is fair.

Disadvantages of the traditional school test. The disadvantages of this type of test are mainly two: first, that the test is nonobjective, especially in its scoring; secondly, that while the range of content may be as great as that of the subject matter which it tests, the test itself is relatively restricted. The first of these disadvantages results in unfair scoring, with its consequent dissatisfaction and the inevitable emphasis which the pupil learns to place upon "passing," regardless of worth, rather than upon measurement and interest in his real achievement; the second may lead either to an overemphasis upon relatively unimportant details of

fact or, if the questions have widespread general significance, to discursive answers extremely difficult to judge. Either of these involve a large amount of the teacher's time for proper analysis.

✧ In scoring this test the subjective judgment of the teacher is constantly needed. This judgment is influenced by many factors other than those directly connected with the evidence on the face of the test, as every teacher knows who has made any attempt to grade such tests. The appearance of the individual, his school record, the general opinion of him held by the examiner, the appearance of the paper itself (its margins, the quality or size of the handwriting, the correctness of the spelling, the style of composition), — these and many other elements tend to influence this subjective judgment. An answer substantially the same given by two different individuals may be given a different grade, whereas two different answers, of intrinsically different merit, may be given the same grade. An answer at the beginning of the reading of a series of papers may receive one grade, whereas the same answer at the end of the reading may receive a different grade. That variations of this kind take place even among the best-qualified examiners, and even in subjects such as arithmetic and geometry, where the procedures and methods seem very clearly given, has been amply demonstrated. In spite of this objective evidence to the contrary, teachers in general are convinced of the fairness of their judgment and of the validity of their scoring. Pupils frequently, therefore, make as great a study of the vagaries of their instructors as of the subjects themselves. In college, in many cases within the writer's observation, the studying for an examination is a guess of the idiosyncrasies of the instructors, and examinations are carefully kept, annotated, and passed on as a heritage to future generations of students.

✧ The relative restrictedness of the traditional school test may be shown graphically as follows: In an average test, say of four questions and with a group of four pupils, let it

be granted that for a certain unit of subject matter each of the four retains a knowledge of 50 per cent of the materials involved, but a different 50 per cent in each case. On this basis, and also on a 100-percentile marking system, each of the four should receive the same mark, namely 50 per cent, since each knows the same amount of the materials. For the sake of argument it is also supposed that these four individuals can put on paper all that they know, and cannot put down on paper what they do not know.

If the four questions are distributed throughout the subject matter in the manner shown in Fig. 2, Question 1 involving for its correct answer a knowledge as indicated by the line shown, Question 2 as indicated by the next line below, and so on, it may be seen that pupil I is able to answer satisfactorily the first and second questions but is unable to answer the third and fourth. This gives him a mark of 50 per cent, which is his true mark.

On the other hand, with the same questions and with an equivalent amount of knowledge of the subject, pupil II is unable to answer any of the four questions and therefore receives a grade of zero, when his actual mark should be 50 per cent. Pupil III is able to answer all the questions, although he knows no more, from an absolute standpoint, than pupils I and II. His grade is, nevertheless, 100 per cent. Pupil IV is able to answer three of the questions and is unable to answer the remaining one. Therefore his grade on this examination is 75 per cent.

Even if there were no difference in the marking of this examination owing to the personal judgment of the examiner, such a possibility as has been shown above means that this

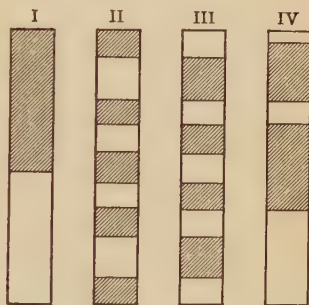


FIG. 1. Shaded portion indicates distribution of knowledge of four pupils

type of test is inadequate to measure fairly a knowledge of subject matter, and is unfair to the pupils.

The development of standard tests. Until the past few years there have been no tests more reliable in their results than these, and particularly no other methods of testing which were available for the classroom teacher. The problem has, however, been appreciated for many years, and

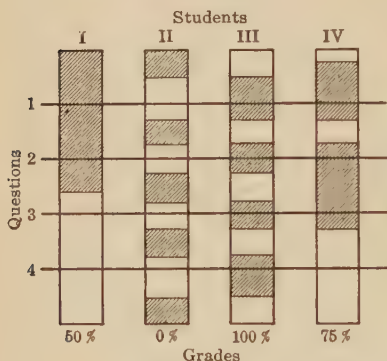


FIG. 2. Possible grading inequalities among four pupils having the same total amount of knowledge

in England, as early as 1864, E. B. Chadwick reported in a magazine article an attempt on the part of the Reverend George Fisher of the Greenwich Hospital School to make standards by means of a "Scale Book."¹

In 1897 Dr. J. M. Rice, in a meeting of the National Education Association, raised the question how much the differing amounts of time used in different school systems in

such subjects as spelling affected the actual final achievements of the pupils. A lively discussion of opinion followed, which stimulated Dr. Rice to initiate an experiment that would show the facts. His experiment consisted of a uniform test in spelling which was given in various school systems and on the basis of which valuable comparisons were made. The comparisons controverted the belief that increased increments of time spent in school upon spelling brought proportional increments of achievement in spelling.

Dr. Rice established the fact that it was both necessary and possible to measure results in education, and the next

¹ Reported by M. R. Trabue, in *Measuring Results in Education*, pp. 70, 71. American Book Company, 1924.

step was to develop the proper methods for so doing. The first step in this was the publication in 1904 by Dr. E. L. Thorndike of Teachers College, Columbia University, of a volume on mental and social measurements, which showed how statistical procedures already in use could be applied to the service of education. A combination of Dr. Thorndike's statistical method with the subject-matter content as suggested by Dr. Rice resulted in the publication in 1908 by Dr. C. W. Stone of the first Standard Test, known as the Stone Reasoning Test in Arithmetic. Dr. Thorndike himself followed, in 1910, with the publication of his Scale for Handwriting, which was followed in 1912 by the Hillegas Scale for the Measurement of English Composition of Young People.

The essentials for standard testing were now supplied. Further refinement, greater use, and further application were necessary, and testing would be placed on a firm footing. The essential value of the Standard Test in school work was early demonstrated by Dr. S. A. Courtis of Detroit, and he devised both new tests and procedures. He was the first to use tests in the measurement of a school system in New York City in the school inquiry of 1912-1913. In 1914 tests were first used in the survey of the Springfield, Illinois, schools to measure the efficiency of a school system. Many other tests were soon made, among them the Buckingham Spelling Scale, the Ayres Spelling Scale, the Thorndike Reading Scale, the Woody Arithmetic Tests, the Woody-McCall Arithmetic Mixed Fundamentals, the Thorndike-McCall Reading Scale, and others. They covered a wide range of the curriculum of the elementary school, spread into the subject matter and range of difficulty of the high schools, and were finally introduced into the work of the colleges. New methods were developed and further tests devised, until at the present time there are available many tests of various sorts for almost all subjects in practically every stage of our educational program, from kindergarten to college.

Another development in standard testing which has had great influence on the work of the classroom teacher has been the development of tests for general ability — the Intelligence Tests. These had their origin in the individual examinations for intelligence devised by Dr. Alfred Binet in France, which were later brought to this country. After revision these Binet tests first appeared here as the Stanford Revision of the Binet-Simon Tests. Since then other revisions have been made for various purposes. These, however, as well as the original test, were designed to measure individual children, as individuals, and not in groups. This fact limited in great degree the scope and adaptability of the tests, and there followed efforts on the part of psychologists to devise a similar scheme adapted to group administration so that a large number of individuals could be measured at one time by a single examiner. The result was a number of group tests for the measurement of intelligence, now made standard and reliable. There was developed a definite procedure, adapted partly from the methods of educational testing and partly from the methods previously used in intelligence testing, which later played its part in the World War in determining the relative abilities of the soldiers located in the large training-camps. Since the war several modified forms of the army tests have been published for use with groups of children of school age; from these the classroom teacher can derive much useful information and much help in teaching.

It was early discovered that the reliability of all these tests depended largely on the closeness and accuracy with which they were given and corrected. Standardization meant that all the children receiving the test should be treated in exactly the same way; that of thousands of children answering a single question with varying answers, all who answered the question in the same way should receive the same credit; and that the amount of credit should be proportional to the difficulty of the question. This was a difficult task, but every

effort was bent toward making it possible. Ambiguities were eliminated as they were discovered. Wording and phrasing were changed to improve the tests. Wrong suggestions were corrected, and the tests were continually altered to meet the needs better as these were revealed. One of the developing characteristics was an effort to make the scoring and interpreting of the results of the tests as easy as possible. This meant making the answers short and pithy, making certain spaces for the recording of the answers to the various parts of the test, listing the various possible answers together with their varying valuations, and in every way mechanizing the tests to improve their use in schools, their ease of administration, and the reliability of their scoring. The tests that are on the market today, known as Standard Tests or Scales, show the evidence of these efforts and the care with which they have been made. Many of them are in the form of printed booklets or folders. On them are found the identification marks, such as the name of the pupil, his school, his grade, his age, and his section; minute and clear directions to guide the teacher in giving the test; equally minute and clear directions for the guidance of the pupils taking the test; frequently a practice form for the elimination of technical difficulties encountered by pupils unfamiliar with the taking of tests; spaces for the answers to the questions; and complete directions for the teacher in scoring and interpreting the scores of the tests. In addition to all this there are furnished the "norms" of the test, which are of great value because they enable teachers who use the tests to make comparisons of the pupils in their own schools with those of the rest of the community, with pupils of any particular section of the nation, or with the pupils of the nation as a whole. In addition, the norms form a real basis for the judgment by a teacher of his own teaching, or a judgment of the degree of efficiency of the work of the pupils. The norms are derived in several ways and are published in several forms. Some norms are derived from results secured from

as many pupils as have been tested in every section of the country. Others are derived from selected samplings of large cities or school units, and still others constitute the norms for a particular section of the country. Many norms are given as "grade" norms. Others are given in terms of "age," and some for both "age" and "grade," which is most useful and desirable.

It would be difficult indeed to put too much emphasis upon the tremendous benefits which classroom teachers have derived or may derive from Standard Tests. They have enlarged the values of teaching and have reduced the errors of examining. They have made comparisons, diagnosis, and remedial treatment possible, as well as the improvement of teaching. The new techniques, such as the Ratio techniques, too, give to teachers a leverage in teaching which has never before been possible, and make analyses of the achievement of children, in terms of their abilities to achieve, a new tool in teaching.

Disadvantages of the Standard Tests. The Standard Tests have several disadvantages from the standpoint of the classroom teacher. The first of these disadvantages is the cost. The tests are sold by different agencies, and although they are desirable and even necessary for all teachers to use in their classes, except for a few times during a year the cost is prohibitive for many teachers. All schools and classes should be able to afford the use, even if a limited use, of the Standard Tests, but it is recognized that all classes and all teachers cannot afford their exclusive use.

A second disadvantage, which is rapidly being corrected and made less in many cases, lies in the fact that a teacher frequently needs measures at shorter intervals than are feasible with the few forms which many tests have. Some tests are published in many alternative forms, it is true, but many tests have only one form or at most a few, and this limits very materially their usefulness to the teacher.

A third disadvantage lies in the scoring difficulties. To

get the greatest value from the tests — in fact, to get any reliable comparative data — they must be scored minutely, and exactly as the directions accompanying the test direct. These directions for certain tests and for many teachers present real difficulties, although the authors and publishers of tests are making constant efforts to simplify scoring. Teachers who use such tests should have the benefit of a good course in the use of Standard Tests, but even then the scoring of papers so as to be truly fair to all the pupils, and also to achieve the most reliable results, is frequently difficult.

A fourth disadvantage of the Standard Tests lies in the difficulties in the interpretation of data. The experienced tester can get from a series of test scores much valuable information that is not seen by the inexperienced user of the tests; and often, although the results of the tests are reliable, it is difficult for some teachers to make the proper deductions from the facts. This, it must be admitted, is not so much a disadvantage of the tests themselves, but rather a limitation which standard testing imposes on teachers inexperienced in their use.

A fifth disadvantage of the tests lies in the fact that in geography, in history, and in other subjects outside of the skills in arithmetic, spelling, composition, and the like, it is difficult for some teachers to find tests which will test the local range of knowledge or of subject matter that has been locally taught. Most of such tests, adapted as they are for universal use, can contain only those elements of general and widespread usage. Because of this, local differences in the curriculum and in the content of the course of study, and local changes from standards which are otherwise quite general, cannot be included in the standardized forms of tests. From one standpoint this fact makes the Standard Tests in these content subjects valuable; but when the teacher wishes to measure the full local achievement of pupils in a satisfactory way, these tests prove somewhat inadequate. They need to be supplemented by other tests.

All in all, the Standard Tests form a tremendously useful and invaluable form of aid to the classroom teacher. Every teacher should be cognizant of their value and be skilled enough in their usage to realize a large proportion of this value; but the factor of cost, the factor of the limitations of the ground which is covered by the tests, and frequently the factor of the limited number of forms in which many tests are published means that teachers should use other forms of testing in addition.

Teacher's Classroom Tests. To supplement the use of the Standard Tests it is necessary for the teacher to devise tests of his own which will give results of a kind which can be used at the times between the giving of the Standard Tests. As was stated earlier, the traditional school examinations are too unreliable and are too rarely scored with accuracy, and the data obtained from them are too limited both in quantity and in quality, to be of wide value in realizing many of the possible benefits of testing. To cover this deficiency there have been devised and put into successful use a number of means of testing which are, from these points of view, much more satisfactory than the traditional tests. These tests are here called Teacher's Classroom Tests, or they might, with equal justice, be called Nonstandard Tests.

Teacher's Classroom Tests are so constructed as to retain as many of the values of Standard Tests as may be, and at the same time to retain certain of the values of the older type of school tests. They are a direct development from the Standard Tests and retain much of their technique. They are, however, made by the teacher, and are adapted for use only in the room and with the class for which they are designed. They do not in any sense replace the Standard Tests in the same field, but merely supplement them and enable the teacher to keep a closer control of the work of his pupils.

Advantages of Teacher's Classroom Tests. These tests have several outstanding advantages. They are inexpensive.

A teacher can devise as many of them as are needed or desired at no more expense than would be required for the older type of teacher-made examination. The extra time which is required for the preparation of the tests is compensated by the saving in time which is gained in the scoring, grading, and interpreting of the papers, and the teacher is enabled to spend time in making a useful test rather than in wasteful reading of discursive answers. The tests are simple to give, as is illustrated in the following chapters, where the tests are described. Dictation, writing on the board, or mimeographing can be used as the conditions or necessities require. The tests are not so objective as the Standard Tests, but they are far more objective than the usual teacher-made test. By eliminating as far as possible any personal feeling or identification of the papers, the careful teacher can come to an impartial judgment of any pupil. The tests have a wide range of usage, since they can be used for any of the purposes which are characteristic of Standard Tests *within* a classroom, and as far as the classroom is concerned the values are almost if not quite as great. Moreover, when they are used in conjunction with the Standard Tests, the values of both types are increased.

Teacher's Classroom Tests can be adapted to the subject matter that is being taught or to any unit or combination of that matter. They can be used for grouping pupils within a class, for grading pupils and marking them, for finding the distribution of the group, or for diagnostic purposes. When scaled (this will be described in a later chapter), scores from different tests can be equated, added, or compared. The tests can be used for testing the efficacy of remedial measures or even be used as remedial devices. They are valuable in the making of reviews, in the organization of subject matter, and as a supplement to the teacher's judgment in the classification and promotion of pupils. One of their most valuable uses has been discovered to be the added vitality which they give to school work, the added interest with which they

inspire pupils, and the increased attentiveness and vigor with which pupils attack the work which they are asked to do. These advantages will be elaborated, and the means for achieving them will be shown, in the succeeding chapters.

How these tests are an improvement over the traditional school tests in the same field may be shown in the following diagram, which gives the same facts with relation to the four pupils as were shown earlier in the chapter, but shows the many additional points that the Teacher's Classroom Tests



FIG. 3. Many contacts give a fairer measure

can touch. The fact that the pupils have very little writing to do means that this test, which covers a far wider range than the traditional test and covers it far more effectively, can be given in just about the same time and with practically no more effort.

Limitations of Teacher's Classroom Tests. It should not be forgotten that these

tests have in themselves several distinct disadvantages which must be appreciated and guarded against by the teacher who uses them. In the first place, and most important, the tests cannot be used to replace the Standard Tests where comparisons are to be made outside of the particular group to which the test is given, and they cannot be used to replace Standard Tests where the ratio techniques based on mental and educational ages are wanted. The tests have no norms, and they are valueless in comparison with results from Standard Tests. The two types are not comparable, because the Standard Test gives an absolute rating as measured against the standards of like age or like attainment over the whole country or for selected standard groups, whereas the Teacher's Classroom Tests that are to be

described give merely a rating relative to the achievement of the other pupils in the class group. Where such a relative rating is of value, as in the work of diagnosis, in marking pupils in subjects for the term, in finding those pupils who are in need of special help or those pupils whose interests should be diverted to other fields, in providing motivation for various types of school study, or in adding more widespread value to the results of Standard Tests, the Teacher's Classroom Tests are of great use.

The tests cannot be used for classification purposes, since the method of classification is one which is based on the norms and standards set up by Standard Tests, although, as has been stated, in a method which includes a teacher's judgment of the relative standing of the pupils the tests can be advantageously used to supplement or rectify that judgment. The teacher should also remember that in the same field, and particularly in the various school skills where there are numerous excellent Standard Tests, the Standard Test, because of its greater accuracy and the greater value of its results, is to be preferred to any test of his own which a teacher might construct.

A caution to the reader. This book has been planned to give to the ambitious classroom teacher a series of tried classroom tests and a method for using them that will furnish, for the groups of pupils with which they are used, as many as possible of the useful results of testing as were described in the preceding chapter. A casual reading of the following chapters, however, may give the reader an impression of arduous effort or even of pedantic processes. A more careful consideration of the processes involved, and a determined effort to master them, in terms always of the results which are desired, should bring a different opinion. The reader is urged to remember that in the main the processes merely require careful practice to bring mastery, and that practice itself will tend to make the work easier as the essential habits are acquired. It is improbable that these habits

are any more difficult of acquisition for the classroom teacher than are long division and percentage for school pupils, or that they are as tedious in the learning. It is also well to remember that the devices shown here are in the main adaptations of devices used in standard testing, and that the teacher will in any event find them of value in increasing his expertness in, and his respect for, Standard Tests.

Chapter summary. The traditional school examination is based so largely on the subjective judgment of the examiner and is usually so narrow in test range that it is difficult to get reliable results. Standard Tests, on the other hand, although of all the tests available they are objective in the greatest degree, are somewhat difficult to score and to interpret, besides being relatively high in cost. Teacher's Classroom Tests retain many of the advantages of the Standard Tests, as well as some of the better qualities of the traditional school examinations. For these reasons they should be used to supplement the Standard Tests, though not to replace them, and teachers who use them should be fully aware of their limitations as well as of their high potential values.

SELECTED BIBLIOGRAPHY

The following give data with respect to the unfairness or inaccuracy of school marks and marking:

MONROE, W. S. *Measuring the Results of Teaching* (chap. i, "The Inaccuracy of Present School Marks"). Houghton Mifflin Company, Boston, 1918.

GREGORY, C. A. *Fundamentals of Educational Measurement* (chap. v (especially pp. 156-158), "Experimental Evidence to show that School Marks are Inadequate"). D. Appleton and Company, New York, 1922.

TRABUE, M. R. *Measuring Results in Education*, pp. 43-56. American Book Company, New York, 1924.

The following give various historical phases of the development of measurement:

TRABUE, M. R. *Measuring Results in Education* (chap. iii, "How Standard Tests Developed"). American Book Company, New York, 1924.

- TERMAN, L. M. *The Measurement of Intelligence*, chaps. i-ii. Houghton Mifflin Company, Boston, 1916.
- PINTNER, R. *Intelligence Testing*, chaps. i-iii. Henry Holt and Company, New York, 1923.

The following gives some of the values of informal tests:

- BUTLER, W. F. "The Value of Informal Tests," pp. 94-119 of *First Year Book*, Department of Elementary School Principals. National Education Association, 1922.

The following are recent publications which give discussions that have been considered in this chapter. They are adapted to levels above the elementary school, but offer a variety of highly suggestive and useful materials.

- BRINKLEY, S. G. *Values of New-Type Examinations in the High School, with Special Reference to History*. Teachers College Bureau of Publications, Columbia University, New York City, 1924.
- PATERSON, D. G. *Preparation and Use of New-Type Examinations*. World Book Company, Yonkers-on-Hudson, 1925.
- RUCH, G. M. *The Improvement of the Written Examination*. Scott, Foresman & Co., Chicago, 1924.

CHAPTER III

THE TRUE-FALSE TEST

Characteristics of the True-False Test. The True-False Test is designed to eliminate the disadvantages of the traditional school examination which result in unfair marking, as noted in the preceding chapter. It consists of a number of statements, some true and some not true, arranged in chance order. The pupils indicate, from the extent of their knowledge of the elements contained in the test, which of the statements are true and which are not.

The True-False Test is a relatively easy test to construct, it is easy to give, it is easy to score, and the results are quickly reached.

Construction of the True-False Test. The first step in the construction of the test is to select, very carefully, the unit of subject matter which it is desired to use as the basis for the test.

STEP 1. SELECTION OF SUBJECT MATTER

In the case here described for illustration the pupils of an eighth grade have been studying the history of the United States and have been especially considering varying phases of periods of the Civil War and industrial expansion. The test is designed to measure a rather general knowledge of this wide period.

STEP 2. CONSTRUCTION OF A SERIES OF TRUE STATEMENTS

The second step in the formation of the True-False Test is to make a series of true statements covering the subject matter concerned. As far as possible these statements should be of such importance and should contain such matter that

they provoke real thinking. Therefore, in their construction it is wise to keep in mind the larger principles which the subject matter contains and to make the statements illustrative of these larger principles. In this way the statements appear to the pupil much as they would occur to him in his life outside of school, and the reactions which the pupil would make to them would or should be somewhat similar. Thus, if the pupil can connect the illustration with the principle on which it is based, he has a real basis for saying that the statement is or is not true. It is probably wise to have at least twenty statements, and a greater number make an even better test when no other type of test is to be given at the same time. After a teacher becomes expert in making the statements, the manufacture of a larger number than twenty and the selection from them of the best twenty is likely to yield better results. For the unit of history which has been mentioned this procedure was followed. A number of the larger principles which were illustrative of the period in question were illustrated in twenty true statements about them. These statements are as follows:

TWENTY TRUE STATEMENTS CONSTRUCTED BY STEP 2¹

1. The attack on Fort Sumter was a good thing for the North, because it united them in thought and deed as they had not been united before.

2. The naval engagements of the Civil War, unlike those of the War of 1812, played little part in defeating the South.

3. Lincoln's greatest task when he became president was to preserve the Union.

4. With the admission of Maine and Missouri into the Union began a struggle between slave and anti-slave movements which ended only with the Civil War.

5. Cotton-raising could be successfully carried on by slaves, because it required few tools and little intellectual skill.

6. Slaves were not owned by all Southern landholders.

¹ This test was constructed and used by Miss Nettie Fehn, Newton School, Toledo.

7. The South opposed the protective tariff because it would benefit but little from it.

8. Constant revision of the tariff duties with the change of political parties hindered business.

9. Texas had her freedom from Mexico almost a decade before she was admitted into the Union.

10. So far only one president of the United States has ever had to stand trial for impeachment.

11. The textile industry created a social revolution which caused a tremendous change in the mode of living of the civilized world.

12. Factories increased the production but decreased the cost of the article produced.

13. American genius for invention accelerated the industrial progress of the United States.

14. The westward movement in the United States was possible in spite of difficult transportation and the lack of communication.

15. By the Free Homestead Act of 1862 Congress encouraged immigration and the Western movement.

16. For a period preceding the Civil War the United States held first place in ocean-carrying trade, but was superseded by England when iron steamships came into general use.

17. The Napoleonic wars interfered with our commerce.

18. The Embargo and Nonintercourse acts failed to protect the trade of the American colonies.

19. The civil-service reform increased the efficiency of the officeholders.

20. A good system of free public schools had been established throughout the North by the beginning of the Civil War.

A few general rules which might be kept in mind by a teacher in the making of these statements will aid in achieving a good test :

1. Make the statement as short as possible, but do not sacrifice clearness for brevity. Telegrams are frequently ambiguous because of their brevity, and the same holds true when a far-reaching statement is expressed in too few words. On the other hand, verbosity also leads to cloudy meanings. The teacher should try to strike a happy medium.

2. Be sparing in the use of dependent clauses and avoid clumsy compound sentences.

3. Do not deliberately write catch questions or questions which can be clearly taken in two ways. The teacher will be surprised often enough with the hidden meanings which pupils will find in supposedly straightforward statements.

4. Express what is wanted in a single sentence, not two or three separate sentences. If it takes two or three separate statements to express an idea, either the idea has not been refined sufficiently for use in the test or else the range of the idea is too wide for a single statement.

5. Make sentences positive rather than negative. This is frequently difficult to do. The sixth statement in the list above is a negative statement. Before it is used in the final test, it should be changed to a positive statement by in some way eliminating the word "not." There is no objection to including a very few such sentences if the ideas they contain are good, and if to eliminate "not" would cause clumsy wording.

STEP 3. ARRANGEMENT IN CHANCE ORDER

The third step in the construction of the True-False Test is to arrange the statements in chance order. In making out the series of true statements teachers will find that one statement may suggest another, and that the finished series has a thread of organization running through it, or through parts of it, by means of which some of the statements are answered by others preceding or following. This tendency is broken by arranging the statements in chance order. If the primary purpose of the testing is to measure the ability of pupils to organize or relate facts, there are other tests, to be described in later chapters, which are better adapted to that purpose than are the True-False Tests here illustrated.

The series of statements given on the previous pages shows what this organization of statements might be. The first

three statements have to do directly with the Civil War; the next three are concerned primarily with phases of the slavery question; the next four statements have a political background; the next three relate to textiles and invention; the next two are about the Western movement of the people; the next three refer to trade and commerce; and only the nineteenth and twentieth statements have little relation to the preceding ones. This thread of organization should be broken, and any system which will arrange the statements according to chance and not according to the judgment of the teacher is satisfactory, because in the end it means that the statements are connected by no design. A good system is to write a series of numbers, in this case from one to twenty, on small cards or slips of paper. The teacher's time will eventually be saved if these cards are made from fairly heavy paper or cardboard. Calling-cards are expensive, but make an excellent pack that can be kept permanently. If the cards are numbered on both sides, time will be saved too. In this connection the numbers 6 and 9 should be distinctive so as to prevent them from being misread when they are upside down.

The cards should be shuffled or thoroughly mixed, and then drawn one at a time. The number of the first card indicates the number of the statement which should be placed first, the number of the second card the statement which should follow, and the remaining statements can be placed by chance in the same way.

STEP 4. CHANCE TRUENESS AND FALSENESS

The fourth step is that of making the statements true or false. It is good policy to do this by chance also, so as to decrease the subjective judgment of the teacher, which is already strong in the mere selection of the statements. This step can well be combined with the preceding step and both be done at the same time. Again any good method for

getting the statements true or false by chance is satisfactory. A simple method is to toss a coin on the selection of each statement, letting heads mean true and tails false throughout the process. It is important that the coin be a perfect one and large enough to toss well. A quarter is good, if not nicked or bent. A coin that is bent, nicked, or hollowed may have a tendency to fall the same side up. It is also important that the coin be really tossed so as to turn rapidly.

Operating as suggested gives the following as the order and truth or falsity of the statements selected above. The first number drawn is No. 20. The twentieth statement is therefore placed first. A coin is then tossed. In this case it comes down heads, and the statement is allowed to remain as it is, because it will be true as it stands. The second number drawn is No. 18, which makes the eighteenth statement become No. 2 in the rearrangement. The coin is tossed and again comes down heads, and the statement remains as it is, because it is true. The third number drawn is No. 13. The coin is tossed and comes down tails, which makes the thirteenth a false statement, and the wording of the statement is changed to make it false. In this case it is a simple matter of changing one word, and the original statement is made false. "American genius for invention accelerated the industrial progress of the United States" becomes "American genius for invention retarded the industrial progress of the United States." The next number drawn is No. 17, which is therefore placed fourth, and remains true as the coin falls heads. The fifth number drawn is No. 19. The nineteenth statement is therefore placed fifth, and it is necessary to make it false, because the coin falls tails. By changing the word "increased" to the word "decreased" No. 19, too, is easily changed to a false statement. This procedure is continued until all the twenty statements have been placed in chance order and, in addition, have all been determined as true or false.

In the making of true statements into false statements

it will help the teacher somewhat if the following suggestions are observed and the statements changed accordingly :

1. The completed statement should be definitely false, but should be positive if possible. This means that "not," "cannot," "no," and negative verb prefixes should be eliminated. Words that in themselves reverse the truth of a statement are undesirable. When both the statement and its answer are negative, ambiguities may result.

2. It is a good plan to write the statement so that the correct answer results logically as "Yes" or "No" to the question (understood) "Is this so?" There is no objection to placing statements in the form of a question, though it makes a test somewhat monotonous when all the statements are in that form.

3. It is unwise to give leads or suggestions that the statement is false. This merely means that the statements should be as natural as possible, as a statement that is patently artificial is misleading.

STEP 5. ELIMINATION OF DIFFICULTIES

The next step is to scan each statement very carefully with a view to changing the wording to make the statements clearer or better and to prevent as far as possible mere "catch" statements or statements that are unfair from the standpoint of what the class has been taught. Statements should be freed from negatives, if such still exist, and especially from ambiguities which would tend to cloud or obscure the meaning. Every effort should be made to make the test fair and not one which is purposely misleading.

In carefully reading over the statements for the test given above, the wording was changed in a few instances.

Statement 1 was inaccurately written, as the pronoun did not agree with its antecedent in number. The wording was changed to read "for the Northern people" instead of "for the North."

Statement 4 seemed somewhat clumsy. It was improved by eliminating the first word, "with," and changing the verb, "began," to "started." The changes made a much smoother statement.

Statement 5 was loosely worded, but it was improved by changing the word "skill" to "ability."

Statement 6, when turned into a false statement, read "Every Southern white man was a slaveholder." This was made clearer by changing the wording to "Every Southern white man owned some slaves."

Statement 10, when changed into a false statement, read "No president of the United States has had to stand trial for being impeached." This was an undesirable wording, but no better was found. The entire statement was changed to read as follows: "Both President Grant and President Johnson stood trial for impeachment." This still contained an error, and the words "for impeachment" were changed to "of impeachment" in the final wording.

Statement 16 was too long and cumbersome; so the wording was changed to read "The United States lost supremacy in ocean-carrying trade to England when iron steamships came into general use."

The statements as finally rewritten and reorganized follow. The number in parentheses indicates the original numbering of the statement. The number that follows indicates the revised numbering according to chance.

REARRANGEMENT OF STATEMENTS IN CHANCE ORDER AND CHANCE TRUENESS

(20) *True*. 1. A good system of free public schools had been established throughout the North by the beginning of the Civil War.

(18) *True*. 2. The Embargo and Nonintercourse acts failed to protect the trade of the American colonies.

(13) *False*. 3. American genius for invention retarded the industrial progress of the United States.

(17) *True*. 4. The Napoleonic wars interfered with our commerce.

(19) *False*. 5. The civil-service reform decreased the efficiency of the officeholders.

(8) *True*. 6. Constant revision of the tariff duties with the change of political parties hindered business.

(1) *True*. 7. The attack on Fort Sumter helped the Northern people, for it united them as they had not been united before.

(2) *False*. 8. The naval engagements of the Civil War, like those of the War of 1812, played the largest part in bringing success to the victors.

(12) *True*. 9. Factories increased the production but decreased the cost of the article produced.

(5) *False*. 10. Cotton-raising could be successfully carried on by slaves, because it required few tools and a high degree of intellectual ability.

(3) *True*. 11. Lincoln's greatest task when he became president was to preserve the Union.

(9) *True*. 12. Texas had her freedom from Mexico almost a decade before she was admitted into the Union.

(15) *True*. 13. By the Free Homestead Act of 1862 Congress encouraged immigration and the Western movement.

(11) *False*. 14. The textile industry created a social revolution which caused little change in the mode of living of the civilized world.

(7) *False*. 15. The South favored the protective tariff because it was so largely benefited by it.

(4) *True*. 16. The admission of Maine and Missouri into the Union started a struggle between slave and anti-slave movements which ended only with the Civil War.

(6) *False*. 17. Every Southern white man owned some slaves.

(10) *False*. 18. Both President Grant and President Johnson stood trial of impeachment.

(16) *True*. 19. The United States lost supremacy in ocean-carrying trade to England when iron steamships came into general use.

(14) *False*. 20. The westward movement in the United States was possible because there was such a splendid system of transportation and communication.

It will be noted that in the arrangement given above there are eleven true and nine false statements. It is not wise to have the same proportion of statements at all times, and in the method which has been suggested the same number would occur only by chance. It is not wise, however, to permit too wide a variation between the number of true and of false statements; so that if the proportion of true and false statements, where there are twenty, differs from nine true and eleven false, or ten true and ten false, or eleven true and nine false, the teacher should correct the proportion by changing some of the statements.

Giving the tests by dictation. After the pupils have become used to taking the True-False Tests, the difficulties involved in giving them and the dangers resulting from the misinterpretation on the part of the pupils of what they have to do are considerably decreased. The first few times that the test is used, therefore, should be rather carefully planned and each step anticipated. In introducing the test the teacher might say :

I am going to try something new today in which you should be interested. Listen carefully and I will tell you what to do, and you should do just as well as you can. We have been studying about [some events in the history of the United States], and I want to see how well you know some of the things we have been talking about.

Take a piece of paper and a pencil.

At this point it would be wise for the teacher to make sure that each pupil is supplied with the same kind of paper, or at least with paper that is the same size and that has ruled lines the same width apart on the page. Attention to this detail will mean a great saving of time at later stages of the test and will make the work of the teacher easier. It is also advisable for the teacher to see that each pupil has at least two sharpened pencils available at the beginning of the test,

so as to avoid wasting time in resharpening broken points. This should be made a habit to be exercised before any test is given. It is not wise to permit pens or fountain pens to be used, as blots are difficult to prevent and interfere with the legibility of the writing. The teacher may then continue:

At the top of the paper write your name. After you have written your name write the date on the next line below. On the next line below that write the words "Eighth-Grade History Paper." Now turn your paper over so that what you have written is on the back.

The purpose of this move is to arrange the work so that when the papers are being corrected, the names of the pupils will not be seen by the examiner and also so that the papers can be easily identified when necessary. After each of the directions the teacher should pause, and not continue until all the pupils are ready. The pupils should be watched carefully to see that they are all following and understand what is wanted. Pupils who show a need of help may be given it, provided that the help given is in the arrangement of the work and not in the answering of the statements.

Have you all written your names on the back? Are all the papers turned over? Now number down the sheet from one to twenty, like this. Leave a margin of one inch on the left-hand side.

The teacher should here hold up a sheet which is correctly numbered, so that all the pupils can easily see how it is to be done, or if more convenient the blackboard may be used as a model form. Here, again, the teacher should pause between each of the directions to give time for them to be carried out.

I shall read you some sentences. Some of these sentences are true and some of them are not true. See if you can tell me which are which. You must think carefully and do your best. You may have plenty of time. If you think a statement is so, write the word "Yes"

after the right number. If you think it is not so, write the word "No." If you do not know, you may guess, and write down what you think it is.

A good plan at this point is for the teacher to write plainly on the blackboard the following:

If the sentence is true write "Yes."

If the sentence is *not* true write "No."

It is unnecessary for a teacher to follow these directions verbatim. If the pupils are encouraged to take the test without necessarily knowing that it is a test, and if they are made familiar with the required steps in taking it, the ends of the introduction are served. After a preliminary somewhat similar to the above, the teacher might continue:

I shall give you a trial sentence. Do not write anything, but listen carefully.

The teacher should then read a short statement, so easy that most of the pupils can recognize its truth, and should follow that with a discussion of its answering.

How many of you think that this is true? Raise your hands. How many of you think that it is not true? Raise your hands.

Unless the statement is obviously true there will be a division of opinion. In this case each group should be asked why it thinks as it does and the proof be presented that the statement is true. When all the pupils are in agreement that the correct answer to have put down would have been "Yes," they are ready to receive the test proper.

Now take your pencils and listen carefully to me. I will read a sentence. Think about it and try to decide whether it is true or not true. I will repeat the sentence, and then I will say, "Write." When I say "Write," you should put down "Yes" if the sentence is so, or "No" if you think it is not so. Are you ready?

The first sentence is, "A good system of free public schools had been established throughout the North by the beginning of the Civil War." Listen carefully again. Number One: "A good system of free public schools had been established throughout the North by the beginning of the Civil War." Write.

At this point the teacher should allow a pause long enough for all the pupils to make their decisions and to place upon the paper the answer which they wish to make.

Are you ready? Listen again carefully. The second sentence is, "The Embargo and Nonintercourse acts failed to protect the trade of the American colonies." I will repeat. Number Two: "The Embargo and Nonintercourse acts failed to protect the trade of the American colonies." Write.

This procedure should be continued until all the twenty statements have been given and the pupils have answered all the questions. Of course, after the pupils have become familiar with the procedure, the greater part of the directions given here can be eliminated, especially the trial sentence and the interpolated directions. However, the teacher should be very careful to continue certain of the directions in all cases, no matter how familiar the pupils may become with the method. This is a case where familiarity is likely to breed carelessness, and here carelessness is fatal. The teacher should always explain before the test the proper symbol for the true statement and the proper symbol for the statement that is not true, and whenever the test is dictated, as above, the symbols should be clearly written on the blackboard. In addition, the teacher should always repeat the statements, and with the statement, each time, repeat its number. This will enable pupils who on occasion leave out certain answers to find themselves and go on without the danger of becoming hopelessly lost.

Criticisms of the dictation method. For the average teacher the dictation method is probably the easiest way of giving the test, though it is not the best as a rule. It involves less preparation than any of the other ways, and in many cases is about as satisfactory. It has the disadvantage, however, of being entirely oral, and as such has its limitations. There are pupils who can judge more intelligently if they have the statement before their eyes and can read it over several times before answering. This disadvantage can be largely eliminated in the following way of giving this test. Dictation has a decided advantage, however, in the fact that it is much the simplest method from the teacher's point of view. The time the teacher spends in giving the test synchronizes with the time the pupils use in writing it. The other methods require more time on the part of the teacher in preparation, though they take less of the class time in writing the test.

Giving the test by the blackboard method. Many school-rooms have large wall maps attached to rollers at the top of the front blackboard. These make convenient curtains which, if pulled down, hide any writing that may be on the board. Failing this expedient, it is possible for a teacher to improvise, with wrapping-paper or cloth, a wall covering of similar nature. The twenty statements of the test may be written on the blackboard, properly numbered, and spaced so that there is no danger of misinterpretation. This writing takes the place of the dictation earlier described.

The pupils receive the same instructions with respect to preparing the numbered sheets of paper as in the dictation method, and the names are written on the backs of the sheets in the same way. The general procedure is somewhat the same. The pupils should be instructed to look at the sentences, *with their numbers*, to decide whether they are true or false, and to indicate the answers on their papers by writing "Yes" if they think the statements are true and "No" if they think they are not true. After a single expe-

rience with this method the pupils will be found to have mastered it, and therefore in future tests the teacher can eliminate most of the preliminary explanations.

The great change in the blackboard method as compared with the dictation method lies in the fact that the test must be timed, which means that all the pupils should have the same opportunity of starting and finishing together. For the first few tests the teacher should allow plenty of time for all pupils to do all that they can ; in later tests he can shorten the time as he sees fit. Frequently a short-time test will give as good results or better (if a distribution is all that is wanted) than a longer-timed test, since the short test also puts a premium on right decisions quickly recognized and made.

Criticisms of the blackboard method. Although this way of giving the test corrects some of the disadvantages of the dictation method, it has, as well, some disadvantages of its own. The values of the test can come only if the pupils do their thinking during the progress of the test and not prior to it. Writing the statements on the blackboard may make it possible for some pupils to get an inkling of what is to come. It introduces an air of mystery which is all right if it is healthy, but which is undesirable if any opportunity is given, knowingly or unknowingly on the part of the teacher, for some of the pupils to see part of the test before it is given to all. It is not easy for a teacher to write a series of statements on a blackboard and at the same time be sure that they are not being read by someone who may be actively interested in knowing just what they are. For this reason a teacher has to be genuinely careful in selecting a proper time for writing the statements and, furthermore, has to take care that the statements are not read by the pupils before the time of the test.

A possibility for the correction of this difficulty lies in writing the statements (as is so frequently done in the more traditional form of school examination) while the test is in

progress. If this can be done without disturbing the pupils, it is satisfactory, but it is unlikely that it can be done at all without considerable disturbance. The mere fact that it takes about as long to answer one of the statements as it takes the teacher to write it means that pupils would tend to follow the writing while it was in progress, and fail to put the proper attention on the work in hand.

It is probably better, all things considered, for a teacher to plan to write the test on the blackboard at some time when the pupils are out of the room, such as at recess, before school in the morning, in a gymnasium period, or in some other period when the pupils are absent. This insures quiet while the test is being taken, and, so far as possible, insures an equal chance for all. It is difficult to do ; but where it can be done, the blackboard method is superior to dictation.

Giving the test by the mimeograph method. There is a third way in which these tests can be given, which both eliminates the difficulties and retains the advantages of the two previously described methods, although for many teachers it presents practically insurmountable difficulties. This may be called the mimeograph method.

Some schools are equipped with devices for copying written or typewritten materials, and where such devices are available this method furnishes a medium for the administration of these tests which is in many respects better than anything that has yet been suggested. All that is needed is a stencil of the test, and the reproduction from that stencil of as many copies as may be needed. The introduction is the same as that in the two methods previously described. The general process should be explained to the pupils, and they should be made familiar with what is expected of them.

It is unnecessary for the pupils to prepare their sheets of paper, as a space provided on the mimeographed sheets will answer the purpose. This has the added advantage of keeping the statement and its answer more constantly and closely connected than it is possible to do otherwise.

A general form for this mimeographed True-False Test is given below. It is well to note a few of its outstanding characteristics.

At the top of the paper, and at the top of each additional sheet, if more than one is needed, should appear its title. It is unnecessary to emphasize the fact that the paper is a test; so it might properly be called a paper, as in this case. The title should also include the grade for which it is designed, as well as the date when it is given.

The space in the left-hand margin may be left for the raw score on this part of the paper, and the space in the right-hand margin may be left for an M score, for use by the pupil in obtaining a grading ratio when that form of grading is used.¹ This allows the corners of the papers to be folded in to conceal the scores while the papers are being returned to the writers. The advantage of this measure will be referred to later in this chapter.

The directions for taking the tests should follow. A simple direction is given in the sample below, which has been used with success. These directions should name the kind of test that is involved and the symbols that are to be used. They should, as well, offer as much encouragement to the pupils as possible.

The body of the test can then follow. The statements should be clearly separated and clearly numbered. In the margin before each numbered statement a space should be outlined in which the answers of the pupils may be recorded, and the directions as to how this is to be done should be included in a preliminary statement at the top of the page.

The True-False Test may cover more than one mimeographed sheet. If so, the second sheet should be clearly numbered and, no matter how many sheets are included, the last item that a pupil can read should be a direction to place his name on the back of each sheet.

¹ See Chapters XIV and XV on "The Making of Composite Test Scores" and "Judging Pupils in Achievement according to Ability."

Sample of Test as prepared for the Mimeograph Method

EIGHTH-GRADE HISTORY PAPER

November 10, 1923

RAW SCORE-----

M SCORE-----

This is a true-false paper. In the spaces in the margin below write the word "Yes" before each statement that you think is TRUE. If you think that the statement is NOT TRUE, write the word "No." Do your best and answer every statement.

- 1. A good system of free public schools had been established throughout the North by the beginning of the Civil War.
- 2. The Embargo and Nonintercourse acts failed to protect the trade of the American colonies.
- 3. American genius for invention retarded the industrial progress of the United States.
- 4. The Napoleonic wars interfered with our commerce.
- 5. The civil-service reform decreased the efficiency of the officeholders.
- 6. Constant revision of the tariff duties with the change of political parties hindered business.
- 7. The attack on Fort Sumter helped the Northern people, for it united them as they had not been united before.
- 8. The naval engagements of the Civil War, like those of the War of 1812, played the largest part in bringing success to the victors.
- 9. Factories increased the production but decreased the cost of the article produced.
- 10. Cotton-raising could be successfully carried on by slaves, because it required few tools and a high degree of intellectual ability.
- 11. Lincoln's greatest task when he became president was to preserve the Union.
- 12. Texas had her freedom from Mexico almost a decade before she was admitted into the Union.
- 13. By the Free Homestead Act of 1862 Congress encouraged immigration and the Western movement.

- 14. The textile industry created a social revolution which caused little change in the mode of living of the civilized world.
- 15. The South favored the protective tariff because it was so largely benefited by it.
- 16. The admission of Maine and Missouri into the Union started a struggle between slave and anti-slave movements which ended only with the Civil War.
- 17. Every Southern white man owned some slaves.
- 18. Both President Grant and President Johnson stood trial of impeachment.
- 19. The United States lost supremacy in ocean-carrying trade to England when iron steamships came into general use.
- 20. The westward movement in the United States was possible because there was such a splendid system of transportation and communication.

Be sure that your name is on the back of each sheet.

By this method, as in the previous one, a specified time should be allowed, and at the end of that time all work should cease. When the sheets are passed out, if there is more than one page of the test, all the pages for one pupil should be clipped together. This will insure each pupil a complete copy of the test and will at the same time prevent confusion during the examining period due to the necessity of passing out further sheets.

The sheets should be placed face down on the desks, and the first directions given by the teacher should include that of having each pupil sign his name on each sheet. If the names, as has been suggested in a previous section, are on the reverse side of the sheet, the pages can be scored without identification. It is important that *both* sheets be signed, as in scoring it is convenient to separate the pages and score all of each kind at the same time.

Before the papers are handed in at the close of the testing period, the pupils should again be cautioned to make sure that their names are signed, as it is the only identification

mark that the sheets contain. Papers cannot be identified by handwriting, because there is so little of it.

The scoring of True-False Tests. The correction of the papers, for which there are several methods, is largely a matter of routine and may be quickly accomplished.

1. *Teacher correction.* One method of scoring may be called "teacher correction," because it involves work done entirely by the teacher.

STEP 1. PREPARATION OF THE SCORING KEY

As the first step in scoring, the teacher should prepare a scoring key. In the three forms of giving the test this scoring key may be the same. For the first two methods the teacher should prepare a sheet similar to that used by the pupils. This should contain the same set of numbers and the correct description of the trueness of the statements following the numbers. Care should be taken that this sheet is spaced in the same way as that used by the class, since the key is to be placed side by side with the papers of the pupils. This is the reason for the caution made above, that all the pupils use the same kind of paper, with identical rulings. In the third method one of the unused mimeographed sheets will be satisfactory. Both on the keys for the dictation and black-board methods and on the key for the mimeograph method the correct markings should be placed in red. In scoring, a red pencil will be found convenient, as the use of the color readily distinguishes the marking of the teacher from that of any of the pupils. After the key has been prepared, all the sheet can be cut away except the strip containing the sentence numbers and their accompanying correct description of "Yes" and "No."

STEP 2. SCORING BY MEANS OF THE KEY

The second step is easy. The teacher places the papers of the pupils in a pile and then slides the master form, or key,

beside that of the answers written on the paper on top of the pile. It is then a simple matter to check those numbers on the pupil's sheet which differ from the key. If the key says "Yes" and the pupil's sheet says "No," a check should be made. The same should be done when the key says "No" and the pupil's sheets say "Yes." When the two answers coincide, as "Yes"—"Yes" or "No"—"No," it is unnecessary to make any mark.

SAMPLE OF SCORING BY MEANS OF KEY

PUPIL'S PAPER			KEY	
Marking of Incorrect Statements	The Pupil's Answers	Numbers on Pupil's Paper	Key Numbers	Key Answers
	Yes	1	1	Yes
X	No	2	2	Yes
X	Yes	3	3	No
	Yes	4	4	Yes
	No	5	5	No
	Yes	6	6	Yes
X	No	7	7	Yes
	No	8	8	No
X	No	9	9	Yes
X	Yes	10	10	No
	Yes	11	11	Yes
X	No	12	12	Yes
	Yes	13	13	Yes
	No	14	14	No
	No	15	15	No
	Yes	16	16	Yes
	No	17	17	No
	No	18	18	No
X	No	19	19	Yes
X	Yes	20	20	No

2. *Pupil correction.* There is another method of correction, which can have as many variations as the teacher is able to devise. This method is to have the corrections made by the pupils themselves. Although it introduces errors in correction (especially when it is first used, for there will be less error after the pupils understand the method of scoring), it has

one outstanding advantage as a teaching-device. It becomes a real teaching-device, because it ties together the mistakes and the reasons for the making of the mistakes, and at the same time emphasizes the correct statements and the reasons which make them correct. While the method has the disadvantage of allowing greater error than the teacher correction, it will usually be found that the pupils are anxious to avoid errors and the consequent criticism of their fellows.

One way of pupil correction consists in having all the papers collected, shuffled, and passed out again to the pupils. The pupils should be instructed to leave the papers lying flat on their desks, as this hides the names of the pupils who wrote them. The next step is for the teacher to read a statement and ask for a class opinion of its correct answer. After a discussion of the question and a decision as to the answer which should have been given, the teacher should state that all who have papers on their desks which have a different answer should mark those statements with a cross, and that all who hold papers which have the same answer should not make any marks. There will be difficulties at first, but the pupils will soon learn how to do the marking acceptably and with little error. This will also be found to be the case even when the pupils mark their own papers, although this is not usually advisable, as it unnecessarily introduces an element of temptation to change decisions.

Pupil correction is on the whole less desirable than is teacher correction, largely because in teacher correction the teacher has an opportunity to get a larger view of the class difficulties than is possible in pupil correction. With this analysis of the errors and the probable reasons for making them, the teacher can make the time of the class in the discussion of the test paper much more valuable than it is possible to do when the pupils correct their own papers. It is usually advisable, moreover, when the pupil scores are to be used for purposes of grading for the teacher to review the pupils' markings as a check.

How to find pupil scores. There are two formulas now in general use which enable a teacher to reach a definite score for each paper. There is some controversy among statisticians as to the absolute fairness of either way of reaching a score, but no plan has yet been proposed which is better. When a new plan of scoring is adopted which is fairer than here proposed, it will be simple for the teacher using the older plan to change to the new. Both formulas now in use presuppose that the pupil knows some statements to be either true or false, and does not know others. This seems perfectly sound. Therefore, if a pupil who knew nothing at all were to take the test, he would stand an even chance of getting half the statements correct and half of them incorrect just from pure guessing. Therefore the formulas further presuppose that he guesses the answers of the true and false statements which he does not know, and by this means is enabled to guess (since there are only two possibilities) some of them correctly and the rest incorrectly. It further supposes that half these guesses will be correct and half of them incorrect. This is not altogether true in the individual case, but it is true that in the long run it will be substantially fair.

The formulas provide as follows: The number of correct answers appearing on a paper is a composite of what the pupil knew to be correct and those that he guessed correctly. But the number that he has wrong represents, on the average, half of what he did not know, since he guessed correctly the other half. The actual number of statements that he knew, therefore, is equal to the number of statements which he answered correctly minus the number of statements which he guessed correctly. The number of statements which he guessed correctly is equal to the number of statements, on the average, which he answered incorrectly. The first formula, then, is as follows:

Formula applicable in all cases:

The number of statements correct minus the number of statements incorrect is equal to the true score.

In symbolic form, where R is the number of right statements, W the number of wrong statements, and S the true score, the formula appears as follows:

$$R - W = S.$$

Let it be supposed that a pupil actually knew ten of the twenty statements given above and actually did not know the other ten. He should have a score of ten, the number he actually knew. However, since he knows ten, he answers all the ten correctly; of the other ten, which he does not know, he guesses five correctly and five incorrectly, and this gives him an apparent total of fifteen correct and only five incorrect, although it is known that this is not his true score. In using the formula given above the true score is found as follows:

$$\begin{aligned} R - W &= S, \\ 15 - 5 &= 10, \end{aligned}$$

where R is 15 and W is 5.

Doing the same thing for the example given on page 48, the substitution is as follows:

$$\begin{aligned} R - W &= S, \\ 12 - 8 &= 4, \end{aligned}$$

where R is 12 and W is 8.

This formula may be used for all True-False Tests, regardless of the number of statements that are attempted, and should always be used when answers to any statements are omitted. Suppose that in the example given above the pupil had done twelve of the twenty statements correctly, seven incorrectly, and had omitted one. In this case the omitted statement would not be counted, and the true score would have been 5 instead of 4. However, when all the statements of the test have been attempted, there is a second formula which shortens the work of converting the corrections into a score. The use of the first formula means that all the correct statements must be counted as well as those

that are incorrect. The second formula means that only the incorrect statements need be counted.

Shorter formula to be used only when all the statements are answered:

When all the statements have been answered, the total number of statements in the test minus twice the number of incorrect statements is equal to the true score.

In symbolic form, where T is the total number of statements in the test, W the number of statements wrong, and S the true score, the formula is as follows:

$$T - 2(W) = S.$$

If this formula were applied in the case first cited above, the substitution would become as follows:

$$\begin{aligned} T - 2W &= S, \\ 20 - 2(5) &= 10, \end{aligned}$$

where T is equal to 20 and W is equal to 5.

In the case cited on page 48 the substitution becomes

$$\begin{aligned} T - 2(W) &= S, \\ 20 - 2(8) &= 4, \end{aligned}$$

where T is 20 and W is 8.

This formula should be used only when all the statements have been attempted and there are no omissions. It will be noted that when so used it gives exactly the same result as if the first formula had been taken.

The tabulation of test scores. The scores on the individual papers are not absolute like those of a percentage scale; they are relative scores, and can be interpreted only in the light of what the class as a whole has done. It is necessary, therefore, for the scores of the entire class to be tabulated before the scores can be interpreted. After the scoring is completed, each paper should be checked to make sure that the score given is correct, and the papers are then ready for tabulation.

The tabulation of the scores of the True-False Test is

similar to that of other tests, described in later chapters. For this reason separate chapters are devoted to a discussion of the ways and means which may be used in that tabulation.

A defense of True-False Tests. An objection occasionally offered to the True-False Test is that it is poor psychology ever to present a false statement to pupils. In the first place, such an objection, when not qualified, merely takes into account one of the laws of learning, — that of Use, — namely, that the exercise of a false bond strengthens the connection. If the teacher allows this to be the case, it is wrong to present such false statements. The Law of Effect in learning, however, is just as potent in strengthening or weakening the bond as is the Law of Exercise; and it is one of the tasks of the teacher to see to it that satisfaction is attached to the right connections and annoyance at failure to the wrong ones, thus strengthening the one and weakening the other. Therefore the tests should always be passed back to the pupils, and each statement should be reviewed by teacher and pupils together so that each may see his mistakes and appreciate why they were mistakes.

In the second place, the truths of life are not always presented in tabloid true form. The roundness of the earth is so little apparent that it was many centuries before its flatness was questioned. We can rarely get a true judgment of a city from its railroad environment. It is the sun which appears to move, and not the earth. A coin in a glass of water seems in a very different place from where it really is.

What is true with respect to the way objects appear to our senses is equally true with respect to our appreciation or interpretation of facts, ideas, or thoughts. Some widely known "facts" are without foundation of truth, yet few people have questioned them. The myth of Washington and the Cherry Tree is widely disseminated; so also is that of the ostrich that hides his head in the sand; and likewise the story of the hoop snake that places his tail in his mouth and rolls away. We rarely question the statements of the

orators of the political party to which we adhere; and we tend to accept with little discrimination the statements of our newspapers.

From this point of view it would seem that we should definitely teach pupils to criticize and judge the truth of statements in terms of the larger principles that we teach.

It may be argued, however, that these objections are not really the most vital, as it is the first impression which a child receives that counts. It seems perfectly reasonable to suppose that a false first impression would be undesirable, but none of the statements on a True-False Test should be "first" impressions. It would not be a fair test of previous teaching unless the matters tested had been previously taught and taught correctly.

Use of True-False Tests for review. After the scores have been tabulated and assembled for use in the classroom, one of the greatest advantages of this test will become clear. This is the use of the True-False Test as a means of review to bring home to every pupil his mistakes and the reason for his making them. The procedure for accomplishing this end may be briefly described.

By consulting the backs of the test-sheet pages the names of the pupils may be found. Thus the papers can be distributed to the various pupils without the separate scores becoming generally known. This can be further helped if the teacher's scores are written in the corners of the papers, and if at the conclusion of the scoring the corners are turned over, inclosing the score number. In this way the score of each pupil becomes his exclusive property, which he can maintain inviolable if he so desires.

When each pupil holds at his desk his own paper, the teacher should read the first statement, questioning the class for their reasons for making the statement true or false. If it becomes obvious that the pupils in general have missed the statement because of some difficulty in the wording or because of some ambiguity which is manifestly unfair, the

teacher should make an adjustment in the scores, especially if these are to be used for M-scale ratios or for marking and grading. In the use of these tests it must be constantly kept in mind by the teacher that it is of utmost importance that the pupils appreciate and expect fairness and justice. Unfairness or injustice must be eliminated wherever possible, because one of the large values in Teacher's Classroom Tests lies in the difference which is created in pupil morale with respect to testing.

When the first statement has been sufficiently explained to make it clear to all the pupils, the next following statement can be given and discussed in like manner. Great care should be taken to emphasize the reasons or principles involved which govern the trueness or falseness of the statement. This will tend to give the pupils the attitude of looking for reasons upon which to base judgments, rather than to stimulate mere blind guessing, and will eventually result in a freer type of thinking and a higher grade of reasoning than would otherwise be possible.

Some who have introduced these True-False Tests in school or college maintain that they are more generally disliked by students than liked, and that the instructor who uses them is accused of doing so to save his time in reading and scoring papers. If for no other reason, this would seem to be a most splendid argument for rather than against their use. However real this dislike may be, it has not occurred generally in the writer's experience either in school or in college, though it has occurred with respect to a few isolated individuals. Where the tester has been willing to accept justified student opinion and to change scoring accordingly, the students have shown themselves more in favor of these and like tests than of the traditional types. On the levels of instruction in the elementary schools, in the writer's experience and that of many grade teachers, principals, and supervisors who have used the tests under his direction, after pupils have become accustomed to the method

of taking the tests, and especially where the results are analyzed by the classes under the leadership of the teachers, the tests have been well liked and preferred to the more traditional forms.

It may be said that the use of the test in this way not only helps the pupils to think but also imposes the same necessity upon the teacher. It is as a result of this thinking and reasoning that pupils find great pleasure in the taking of the tests. It has a large amount of the game element in it, which is stimulating to effort; but while in most games the interest is largely in rivalry with others and the achievements of others, it can in this case be easily turned to that higher type of rivalry — rivalry of one's previous best efforts. It has been found that for some pupils the use of these tests has provided a motivation to real intellectual efforts, even after practically every other means at hand had failed. This was probably due to the fact that there is in this type of test no attempt at sugar-coating, which is so easily detected and so generally resented.

Chapter summary. The success of the True-False Test involves the very careful selection of the unit of subject matter to be used and the construction, on the basis of that matter, of a number of true statements. These statements are then arranged in chance order and are made true or not true on the basis of chance. The final step in making the test includes the careful scanning of the statements as finally arranged so that unfair and ambiguous sentences can be eliminated.

The giving of the test involves at first a detailed introduction to the pupils, which can be eliminated as they become familiar with the method. There are three ways of giving the test: by dictation, by the blackboard, or by mimeograph. Each is good, especially the last if teachers have access to the required apparatus for making stencils and copies.

The test may be scored either by the teacher or by the

pupils; but when the best results are desired, the former is undoubtedly the better method. In either case the formulas used in the scoring, though perhaps unfair to a few pupils, are the same. These formulas are $R - W = S$ when answers have been omitted or whenever the teacher prefers to use it, and $T - 2(W) = S$ when all statements have been answered.

The great value in the use of the tests lies in the review which they afford, and the teacher who uses them in this way will soon appreciate the results. On the upper levels of instruction a too great dogmatism on the part of the instructor or a general misunderstanding of the purposes of the testing on the part of the students has been known to produce among the students a dislike for this type of test, though in elementary schools they are generally well liked and welcomed.

Samples of True-False Tests¹

THIRD-GRADE GEOGRAPHY PAPER²

RAW SCORE _____ DATE _____ M SCORE _____

This is a True-False paper. In the spaces in the margin below write "Yes" before each sentence you think is RIGHT, and "No" before each sentence you think is NOT RIGHT. Do just as well as you can.

- 1. At the slaughterhouses animals are killed and the meat is prepared for the meat markets.
- 2. Bananas, dates, and sugar cane grow in hot countries.
- 3. Range animals are branded so the ranchmen can tell to whom they belong.
- 4. Truck farms are located far from the city.
- 5. A dairy farmer raises cows just in order to sell cheese.
- 6. Silkworms live on oak leaves.
- 7. Chicago is on the northern part of Lake Michigan.
- 8. Silk raisers allow all the worms to turn into moths.

¹ These three samples are tests in the third, fourth, and fifth grades in geography.

² This test was constructed and used by Miss Bernice Raymond, Harvard School, Toledo.

- 9. Chicago is one of the largest meat-packing centers in the United States.
- 10. The silkworm is thirty-two days old when full grown.
- 11. Pasteurizing milk kills half the germs.
- 12. Salt is found almost everywhere.
- 13. Meat is sent to our homes from slaughterhouses.
- 14. When full grown the silkworm caterpillar is an inch long.
- 15. Most of the big stock farms are located in the eastern part of the United States.
- 16. Chickens are raised so that the farmer can sell eggs.
- 17. New York is the chief state for making salt.
- 18. Salt is found in mines, wells, springs, and seas.
- 19. The silk moths are a creamy-white color.
- 20. If hens are given great care and good food they will lay about twice a month.

Be sure that your name is on the back of this sheet.

FOURTH-GRADE GEOGRAPHY PAPER ¹

RAW SCORE----- DATE----- M SCORE-----

This is a True-False paper. In the spaces in the margin below write the word "Yes" in front of each statement that is TRUE and "No" in front of each statement that is NOT TRUE. Do your best and answer every statement.

- 1. In the Far North snow is on the ground only about six months of the year.
- 2. Iceland is larger than Greenland.
- 3. The sunshine is very bright near the poles during the winter.
- 4. Greenland has a very warm climate.
- 5. Eskimos build their homes near the water.
- 6. Eskimos make all their own clothes, shelters, tools, and playthings.
- 7. Many of our valuable furs come from northern Canada and Alaska.
- 8. Recently three explorers safely made an exploring trip to Wrangell Island.

¹ This test was constructed and used by Miss F. Drew, Monroe School, Toledo.

- 9. The cold regions of the earth have a sparse population.
- 10. Eskimos live by hunting and fishing.
- 11. Sealskin is a very valuable fur.
- 12. The weather at the south pole is warmer than at the north pole.
- 13. The United States navy dirigible *Shenandoah* may make an exploring trip to the south pole this summer.
- 14. The Eskimo uses dogs to pull his heavy loads.
- 15. The Eskimo of the Far North lives in a stone hut in winter.
- 16. In winter the Eskimos live almost entirely on meat.
- 17. In summer the Eskimo lives in a sealskin tent, or a hut built of stones and dirt.
- 18. Eskimos heat their houses by burning logs of wood.
- 19. Greenland is northeast of North America.
- 20. The sea is so full of ice in the Far North that ships cannot get through.
- 21. Alaska is situated in the southeast part of North America.
- 22. Much agriculture is carried on in the cold regions of the north.
- 23. The reindeer is a useful animal for the Eskimos of Alaska.
- 24. Eskimos are dark-skinned people.

Be sure that your name is on the back of each sheet.

FIFTH-GRADE GEOGRAPHY PAPER ¹

RAW SCORE _____ DATE _____ M SCORE _____

This is a True-False paper. In the spaces in the margin below write the word "Yes" before each statement that you think is TRUE and "No" before each statement that you think is NOT TRUE. Do your best and answer every statement.

- 1. The Chinese invented printing and gunpowder.
- 2. China and her provinces are about the size of Ohio.
- 3. China has fewer people than North America.

¹ This test was constructed and used by Miss Edna Roemer, Auburndale School, Toledo.

- 4. There are many forests in China.
- 5. There is now a railroad from Peking to Paris.
- 6. We carry on less trade with China than we did years ago.
- 7. In China the women as well as the men can vote.
- 8. The Chinese nation is the oldest in the world.
- 9. China has a republican form of government.
- 10. There is less coal in China than in France.
- 11. Most of the people live in China proper.
- 12. The Chinese seldom make things from bamboo.
- 13. Wheat is the greatest crop grown on the plain of China.
- 14. For a long time the Chinese would not let strangers travel in their country.
- 15. Little tea is grown in the United States, because it is hard to grow it here.
- 16. The Chinese eat with chopsticks.
- 17. Only a few people in a tea garden pick tea.
- 18. The tea farms are very large.
- 19. There is plenty of moisture on the plains of China to grow crops.
- 20. There are very few canals or rivers in China.

Be sure that your name is on the back of this sheet.

The following are four interesting examples of the use of the True-False Test in other subjects and for other grades. The first is a test of a part of seventh-grade arithmetic, the second is a test in sixth-grade English, the third is a test for health-teaching in the fourth grade, and the fourth is a nature-study test in the third grade.

SEVENTH-GRADE ARITHMETIC PAPER ¹

RAW SCORE----- DATE----- M SCORE-----

This is a True-False paper. In the spaces in the margin below write the word "Yes" before each statement that you think is TRUE and "No" before each statement that you think is NOT TRUE. Think. Do your best. Answer every question.

¹ This test was constructed and used by Miss Daisy Van Noorden, Lincoln School, Toledo.

- 1. If you know the volume and altitude of a cylinder, you can find the area of one base by dividing the volume by the altitude.
- 2. The radius of a circle with an area of $3\frac{1}{2}$ sq. ft. is 1 ft.
- 3. If the radius of a circle is measured in inches, the area will be inches.
- 4. The volume of a cylinder equals two times the altitude times the area of the base.
- 5. The diameter of a circle is a line drawn from the circumference across the circle to the circumference.
- 6. The area of a circle equals $22/7$ times the radius.
- 7. The area of a circle with a radius of 2 ft. equals 15 sq. ft.
- 8. The altitude, or height, of a cylinder is the distance from one of its circular bases to the other.
- 9. A wheel is circular.
- 10. If you know the circumference of a circle you can find the length of the diameter by dividing the circumference by $22/7$.
- 11. The area of the base of a cylinder is the area of one of its circular ends.
- 12. Any straight line from the center to the circumference is a radius of a circle.
- 13. The volume, or cubic capacity, of a cylinder is the number of cubic inches, cubic feet, or cubic yards, etc. which it occupies or holds.
- 14. The surface of a cylinder is the area of the base times the altitude.
- 15. If you have the radius of a circle given, to find the circumference use $2 \times \text{radius} \times 22/7$.
- 16. The circumference of any circle is equal to $22/7 \times \text{radius}$.
- 17. The perimeter is another name for the circumference of a circle.
- 18. The diameter of a circle is one half as long as the radius.
- 19. The curved surface of the cylinder equals circumference times height.
- 20. A circle is a surface bounded by a curved line called the circumference, nearly every point of which is at the same distance from the center.

Be sure that your name is on the back of each sheet.

SIXTH-GRADE ENGLISH PAPER ¹

RAW SCORE----- DATE----- M SCORE-----

This is a True-False paper. In the spaces in the margin below write the word "Yes" before each statement that you think is CORRECT and the word "No" before each statement that you think is INCORRECT. Do your best and answer every statement.

- 1. He did his work well.
- 2. They seen the child coming down the hill.
- 3. The man had not spoke loudly enough to be heard.
- 4. We had begun our work on time.
- 5. The children came too early.
- 6. They had already eaten their lunch.
- 7. At the picnic the boy had tore his coat.
- 8. The boys had went before the girls.
- 9. The child has run to the store.
- 10. Most of the pupils had their lessons wrote.
- 11. Because they had drunk impure water the children became ill.
- 12. His voice rang clearly across the yard.
- 13. They had sang many of their old songs.
- 14. The man has came too early.
- 15. It was so cold that the river froze.
- 16. This group done its work in the wrong way.
- 17. The man lay very still on the bed.
- 18. We have laid the papers on the desk.
- 19. The child has lain on the damp ground.
- 20. I have gave him my umbrella.

Be sure to write your name on the back of this sheet.

FOURTH-GRADE PAPER ON HEALTH HABITS ²

RAW SCORE----- DATE----- M SCORE-----

This is a True-False paper. In the spaces in the margin below write the word "Yes" if you think the statements are TRUE, and write

¹ This test was constructed and used by Miss Laura Kuhr, Newton School, Toledo.

² This test was constructed and used by Miss Rose Clippinger, Jefferson School, Toledo.

the word "No" if you think the statements are NOT TRUE. Do your best and answer every statement.

- 1. Every person has two sets of teeth.
- 2. A baby's teeth are all right without care.
- 3. The teeth are covered with a hard material called enamel.
- 4. Germs, or bacteria, make holes for themselves in the enamel of the teeth.
- 5. Teeth need exercise to keep them strong.
- 6. The hands should be kept clean, because they may carry germs to the mouth.
- 7. After teeth begin to decay, the dentist can do little to help them.
- 8. Most bubble drinking-fountains are dangerous places to drink.
- 9. The teeth should be cleaned by brushing down on the lower teeth and up on the upper teeth.
- 10. There are thirty-two teeth in the second set.
- 11. The chief business of teeth is crushing and grinding food.
- 12. Bacteria, or germs, grow best in the mouth, because they find both warmth and moisture there.
- 13. If the teeth are not cleaned after each meal, the germs find plenty of food to grow upon.
- 14. Nature provides a young child with a set of sixteen teeth.
- 15. Teeth should be cleaned once each day.
- 16. Cracking nuts, biting thread, and picking the teeth with a pin may crack the enamel.
- 17. All the teeth in our mouths are shaped the same.
- 18. Teeth are of little importance in keeping our bodies well and strong.
- 19. We should visit the dentist once each year.
- 20. Cracks in the enamel of the teeth give germs a chance to get in and grow.

Be sure that your name is on the back of each sheet.

THIRD-GRADE NATURE-STUDY PAPER ¹

RAW SCORE----- DATE----- M SCORE-----

This is a True-False paper. Write "Yes" in front of each sentence you think is TRUE and "No" in front of each one that is NOT TRUE. Think carefully and answer every one.

- 1. All male woodpeckers have red on their heads.
- 2. The woodpeckers sing a sweet song.
- 3. Chickadees like to hunt over the smaller twigs.
- 4. The cardinal is a little smaller than the robin.
- 5. The chickadee is smaller than the English sparrow.
- 6. Chickadees eat many insect eggs from fruit trees.
- 7. The baby male cardinals get their bright colors the second summer.
- 8. The colors of the female cardinal are red and black.
- 9. The red-headed woodpecker is more useful than the downy woodpecker.
- 10. The nuthatch hurts our trees in the winter.
- 11. The downy woodpecker braces himself with his tail when he sits on trees.
- 12. The downy woodpecker is larger than the hairy woodpecker.
- 13. The nuthatch has a black cap on his head.
- 14. Orchards that have chickadees living in them have fewer insects than other orchards.
- 15. The cardinal sings a beautiful song.
- 16. The nuthatch has a short beak.
- 17. The chickadee lives with us only in the summer.
- 18. The downy woodpecker has a black cap on his head and a black bib under his chin.
- 19. Cardinals often raise two families in one year.
- 20. The downy woodpecker has two toes in front and two turned back.
- 21. Cardinals are often called drummers.
- 22. Nuthatches hunt on the trunks and larger branches of trees.
- 23. The cardinal's beak is long and thin.
- 24. The cardinal is a cruel husband and father.

¹ This test was constructed and used (by dictation) by Miss Marie Lerche, Sherman School, Toledo.

- 25. The downy woodpecker likes to run down a tree head-first.
- 26. The chickadee never sings.
- 27. The chickadees like to raise a large family.
- 28. The chickadee eats the eggs only from the top side of the twig.
- 29. The downy and hairy woodpeckers have short, stiff tongues.
- 30. The chickadee's beak is very long.

Be sure that your name is on the back of each sheet.

SELECTED BIBLIOGRAPHY

- MCCALL, W. A. "A New Kind of School Examination," in *Journal of Educational Research*, Vol. I, pp. 33-46, January, 1920.
- MCCALL, W. A. *How to Measure in Education*, pp. 119-133. The Macmillan Company, New York, 1923.
- HAHN, H. H. "A Criticism of Tests requiring Alternative Responses," in *Journal of Educational Research*, Vol. VI, pp. 235-240.
- ODELL, C. W. "Another Criticism of Tests requiring Alternative Responses," in *Journal of Educational Research*, Vol. VII, pp. 326-330.
- BARTHELMESS, H. M. "Reply to a Criticism of Tests requiring Alternative Responses," in *Journal of Educational Research*, Vol. IX, pp. 234-240.
- WEST, P. V. "A Critical Study of the Right Minus Wrong Method," in *Journal of Educational Research*, Vol. VIII, pp. 1-9.
- MILLER, G. F. "A Variation of the 'True and False' Achievement Test," in *School and Society*, Vol. XX, No. 504 (August 23, 1924), pp. 250-251.
- RUCH, G. M. *The Improvement of the Written Examination*, pp. 125-130 and 107-111. Scott, Foresman & Co., Chicago, 1924.

CHAPTER IV

THE JUDGMENT TEST

The Judgment Test involves principles rather than facts. Encyclopedic knowledge, of itself, does not indicate any ability on the part of the holder to use that knowledge for any purpose. Neither does it imply that because of that knowledge there is any underlying familiarity with the principles upon which it rests. If this were the case, as it is not, the work of our schools would be considerably simplified; and it would mean that the art of teaching could safely confine itself, as has sometimes happened, to the imparting of facts, resting secure in the faith that the facts would create their own usefulness. Facts in and of themselves do not create their own usability. It is manifestly impossible for the schools to present all past, present, and future facts in any case, and in view of the famous experiments in the field of memory reported by Thorndike from Ebbinghaus, Swift, Book, and others, it is useless.¹ The schools cannot expect merely to teach facts in the hope that as facts they will be remembered and recalled at will. But the school can teach the facts in relation to the principles which they illustrate, and rest secure in the faith that if the principles are understood through the medium of the facts, the residue of education will be found in the ability of the pupil later to trace from previously unknown facts the principles from which they were derived, and thereby have a just basis for action upon the facts.

This, the ability to recognize the principles upon which

¹ E. L. Thorndike, *Educational Psychology*, Vol. II, chap. x, "The Psychology of Learning." Teachers College, Columbia University, New York City, 1913.

certain facts are based, therefore becomes an important asset for pupils and one which a teacher should encourage and, if possible, measure. It may be that in isolated cases the traditional school examination is able to test such an ability; but if so, it is more by chance than design, and in isolated cases only. In the True-False Test such an ability is of very great help to a pupil in making proper deductions for his answers, and, as indicated at the end of the preceding chapter, the True-False Test can be used to promote this type of logical thinking. It can promote it; it cannot measure it. If the pupil makes a wrong answer to a true-false question, it is evident either that he does not know the principle upon which that statement rests or else that he is unable to connect the obvious statement with the less obvious principle. This is a measure of a kind. On the other hand, if the pupil makes a correct answer, it does not mean that the individual knows the underlying principle which postulates that correctness, because the correctness of the answer may be a matter of chance as a result of guessing. The Judgment Test has been devised to give the teacher a measure that is superior to that given by either of the types of tests described above.

Characteristics of the Judgment Test. In substance the Judgment Test presents to the pupil a number of facts, all of which are true, and asks the pupil to give the reasons which make them true. It is more difficult in many respects than the True-False Test, more difficult for the pupils, and somewhat more difficult for the teacher to score. It is somewhat less objective and is therefore somewhat unfair. Its virtue lies, however, in the fact that in so far as it is unfair, it is unfair to all in the same degree; that in so far as it involves the subjective judgment of the teacher, that subjective judgment is constant for all the pupils.

Construction of the Judgment Test. There are two ways of constructing this test, the differences depending upon the point of view of the maker.

STEP 1. SELECTION OF SUBJECT MATTER

In both methods, as in the True-False Test, the first consideration is to select carefully the extent and character of the subject matter which it is designed to use as the basis for the test. This must be sharply delimited and as clear to the pupils as to the teacher, a unit which has been taught as a unit. Much of the value of this test lies in the attention of the teacher to this first fundamental fact, as will be clearly demonstrated to the teacher who tries the test where this principle has been neglected. The purpose of this is to focus or delimit the attention of the pupils to the subject matter in hand, at least until they have learned how to take the test and appreciate clearly what is wanted. After the pupils have acquired the requisite familiarity and know wherein their efforts are desired, this principle, although of value, is not so pressing, since the pupils will in general limit their answers to the type desired.

STEP 2. METHOD OF SELECTION OF PRINCIPLES, OR METHOD OF SELECTION OF FACTS

After the selection of the unit of subject matter is made, the two methods differ. By the first method the teacher makes a selection of the larger truths or principles which the subject matter has contained. As will be seen, this is an *a priori* method of selecting the *answers* to the questions before making out the questions. After making out these principles and setting them down clearly, the next step is to construct, using the principles as a base, a number of true statements which illustrate them. From this point on, the two methods are again the same.

The second method does not select the principles. The teacher merely makes a list of the more significant true statements that the school work in the subject matter has brought out. These statements are carefully written so as

to eliminate ambiguities, false statements, negative statements wherever possible, and catchy wording. This is done in both methods.

STEP 3. CONSTRUCTION OF STATEMENTS

The next step is to prepare these statements for the actual test according to the method which is to be used for administering the test. Of the two methods described, the second will probably be found to be more useful as well as the more economical in both time and energy. By the first method the teacher is likely to construct statements that read and seem artificial to the pupils. With the principle in mind the tendency is to strip the illustration, so far as may be, of all extraneous ideas, and the result, although it probably is a true statement, is unnatural. By the second method the teacher is more likely to make a series of statements that seem natural and real, even though in some cases they may more than illustrate a given principle. The effort of the teacher in this case may be concentrated on eliminating, by rewording or by changing the sense, any apparent ambiguity. This will usually be found to be an easier and probably more satisfactory way of making the statements. Because of these facts the second method only will be fully described and illustrated.

The Judgment Test given below was constructed on a unit of subject matter which involved the study, in a fourth grade, of the rubber industry in South America. The test is peculiarly well adapted for use with subject matter of this character, as it is fairly easy to construct and reveals another type of ability than that of the True-False Test. There is, of course, a certain necessary literacy required, but this may have only a small bearing, as is illustrated in the sample papers which are given below.

This test was constructed by the second method, and after the selection of the subject matter the next step was the

arbitrary manufacture of a series of statements covering large truths. These were in the nature of facts rather than a statement of principles, and it was hoped that the answers would show a familiarity with the principles which were involved.

SAMPLE OF SEVEN STATEMENTS CONSTRUCTED FOR A JUDGMENT TEST

1. The rubber gatherers sometimes become blind.
2. In a few weeks the Indians must find fresh trees.
3. The Amazon River is a dangerous river to travel on.
4. At night the rubber gatherers wrap up in mosquito netting.
5. Very few white people go to the Jungle.
6. The rubber merchants are Portuguese men.
7. When the rains come the rubber season ends.

STEP 4. ARRANGEMENT OF STATEMENTS IN CHANCE ORDER

The next step is to arrange the statements in chance order. This may be done in any way, but that suggested in the preceding chapter is as effective as any. The process merely involves the writing of the seven numbers on slips of paper or on cards, thoroughly shuffling them, and drawing them in rotation. As the numbers are drawn, the sentence of the number first drawn is given first place, the sentence of the second number drawn is given second place, and the drawing is continued until all the numbers are placed. By this procedure the sentence numbers of the illustration given above were drawn and assembled, with the following result, which represents one chance distribution. The numbers in parentheses indicate the original numbering of the statements when first constructed. The other numbers indicate the new numbering.

SAMPLE REARRANGEMENT OF STATEMENTS BY CHANCE

- (6) 1. The rubber merchants are Portuguese men.
- (4) 2. At night the rubber gatherers wrap up in mosquito netting.

- (2) 3. In a few weeks the Indians must find fresh trees.
- (3) 4. The Amazon River is a dangerous river to travel on.
- (1) 5. The rubber gatherers sometimes become blind.
- (7) 6. When the rains come the rubber season ends.
- (5) 7. Very few white people go to the Jungle.

STEP 5. PLACING THE TEST IN FINAL FORM

With the completion of the previous step the test is ready for placing in its final form according to the particular method which is to be used for giving it. The methods used may be similar to those already described, — dictation, writing on the blackboard, or mimeographing, — and for this test any of the three is practically of equal value, although until pupils are familiar with the technique of taking the test better results will probably come if they have a chance to see and re-read the sentences, as is possible in either of the two latter methods.

Giving the test by dictation. By the dictation method the teacher should see that all the pupils are at their desks and are provided with sheets of paper and sharpened pencils. The test may be introduced in any way that the teacher sees fit, so long as the main object of the introduction is achieved. This is twofold: on the one hand to gain the attention of the pupils, and on the other hand to bring about a willingness and a readiness to take the test. As in the test previously described it is wise to have the pupils write their names and any other wanted information on their sheets and turn the sheets over on the desks, face down, when completed.

If there are seven statements, as there are in this test, the pupils should be instructed to number down the left-hand margin of the papers from 1 through 7, skipping two or three lines between each number. This will allow a space of three or four lines for each answer. To make this clear it has been found desirable for the teacher to illustrate the

various steps on the blackboard. If it seems easier, however, the pupils may be told to number the answers as they are given.

The pupils should then be ready for the test itself. The teacher might say something like the following, although it should be remembered that these directions are merely suggestive and should be changed to suit the individual requirements :

I want to see how well you can answer these statements. I shall read a number of sentences which you know are true and which are about the geography we have been studying. Do not copy the statements. Just write, as clearly as you can, the kind of answer that I will explain to you. Tell me just why each statement is true. I will give you plenty of time to think out what you want to say. Now listen carefully while I read. Number One: "The rubber merchants are Portuguese men." What is the best reason you can give as to why that is true? I will repeat. Number One: "The rubber merchants are Portuguese men."

The teacher should now pause so as to allow all the pupils to think and to write their answers. At first this pause should be long enough for even the slowest pupils to make some sort of decision and to do the necessary writing. Later, when the pupils are more familiar with the method, less time can safely be allowed. The teacher might continue:

Are you all ready? I shall read the second statement. Number Two: "At night the rubber gatherers wrap up in mosquito netting." What is the best reason you can give as to why that is true? Number Two: "At night the rubber gatherers wrap up in mosquito netting."

Here, again, the teacher should pause until the pupils have had ample opportunity to formulate and write their answers, after which the other sentences can be read and

answered in the same way. It will be unnecessary after a short time for the teacher to interject the questions given above, as all that will be needed will be the sentence number and the repetition of the statement.

Giving the test by the blackboard method. By the blackboard method the statements are written on the blackboard as was described in the preceding chapter. The pupils prepare their answer sheets in the same way as for answering from dictation, and the teacher may introduce the test orally in the same way, making only such changes as are called for by the difference in method.

Let it again be urged that great care be taken by the teacher in concealing the statements until the time that the test is to be taken. It is unwise for the teacher to write the statements on the board during the progress of the test. However, should there be a chance for a few pupils to see and read the statements before the written answers are to be made, it may be a good and desirable thing to do. Moreover, if the pupils become used to having the teacher write statements while they are working on others, then writing the statements while the pupils are busy is much the better way.

When this method is used the teacher should set a time limit for the entire group of statements, at the conclusion of which all work should cease, pencils should be placed on the desks, and the papers should be turned over. The early tests should have ample time allowed for all students to finish the tests comfortably, but with increasing familiarity as valuable results will be reached if the time is shortened.

The main disadvantage of both the dictation method and the blackboard method is that there is opportunity afforded for allowing the pupils to mix the numbers of the statements and the numbers of the answers. When such cases occur it would be wise for the teacher, in correcting, to make allowances for the mistakes and give credit for correct answers that are misplaced. Although it takes a little more class time during the progress of the test and introduces a slight dis-

traction, the suggestion made above — to number the statements at the time the questions are answered — will tend to correct this difficulty.

Giving the test by the mimeograph method. The most satisfactory method, when it is possible to use it, is the mimeograph method, as there is no possibility that the pupils will make mistakes in mixing the numbers. In addition, this way of giving the test has the added advantage of directly connecting the original statements and the original answers when the papers are handed back to the pupils for the judging of mistakes and the reasons for making them.

A suggested form for the mimeograph Judgment Test is given below. It will be noted that in this form, as in the form illustrated for the True-False Test, it is labeled a "paper" rather than a "test." The reason for this is that pupils may have had undesirable associations with the word "test" which it would be better to avoid in this newer form of examination. Using the word "paper" may not correct these associations, it is true, but it may do so, especially after the pupils learn the difference between these tests and those that they have been in the habit of taking. It should also be noted that the space for the raw score is on the left-hand side, near the corner of the paper, so that it can easily be turned over to conceal the mark, and the space for the M-scale rating is in the opposite upper right-hand corner.

The directions here given have been used with success, though many variations are possible. Some teachers in the lower grades prefer to use the word "story" to "statement." Whatever word is most familiar to the pupils should be used, since clear and uniform intent is the main purpose of the directions.

A direction to remind the pupil to place his name on the back of each test sheet should always be included as the last element in every test, and if the reminder is placed as indicated, it will save the teacher much time in the identification of the writers of papers.

Sample of Judgment Paper by Mimeograph Method

FOURTH GRADE GEOGRAPHY PAPER ¹

RAW SCORE DATE M SCORE

This is a Judgment paper. In the blank lines below write a short sentence which will tell the best reason you know why each of the statements is true. Do your best. Answer every statement you can. Write clearly.

1. The rubber merchants are Portuguese men.

Answer

2. At night the rubber gatherers wrap up in mosquito netting.

Answer

3. In a few weeks the Indians must find fresh trees.

Answer

4. The Amazon River is a dangerous river to travel on.

Answer

5. The rubber gatherers sometimes become blind.

Answer

6. When the rains come the rubber season ends.

Answer

¹ This test was constructed and used by Miss Beatrice E. Mathias, Pickett School, Toledo.

7. Very few white people go to the Jungle.

Answer -----

Be sure that your name is on the back of each sheet.

Scoring the answers to Judgment Tests. The scoring of the Judgment Test is quite different from that of the True-False Test and is somewhat more difficult. The answer sheets should be collected and piled in front of the scorer, with space so that the sheets can be conveniently turned over and repiled. The answers should be scored separately, all answers to the first statement being scored on all sheets before scoring any of the answers to Statement 2, and all answers to Statement 2 being scored before any to Statement 3 are attempted. In this way the teacher can go through the papers rapidly, keeping in mind the varying answers and their appropriate scores. A good method to use is the following :

Beginning with Statement 1 the teacher should sample about half the answers rapidly so as to get a general idea of the variety of answers that have been given. These may be jotted down on a slip of paper as they are found, and the teacher should make allowances for differences of wording for the same idea. In the test illustrated above the varying ideas, as expressed in the reasons given in answer to the first statement, were as follows :

Live in Pará.

Know language.

They are not English.

They are English.

They gather rubber.

They can stand the heat.

They can stand the heat and know the language.

They live in the Jungle.

Because of weather.

Can stand black smoke of palm leaves and nuts.

It can be easily seen that there is a wide variation in the merit of these answers, although each of them is in some way reminiscent of the study of the rubber industry of South America. The answers, such as these above, should therefore be listed in the order of their merit, with the best answers first and the poorest answers last. A fixed score should be assigned to the best answer; for example, 3. Lesser scores of 2, 1, and zero should be assigned to the less worthy answers. It may be found that two answers are of equally good or of equally poor merit, in which case they should be allowed the same credit. It may also be found that no one has an answer that is worthy of the highest mark; in that case the teacher should conclude that either there is some difficulty in the wording of the question itself, so that it does not convey the meaning that was intended, or else the idea called for was not stressed sufficiently in the teaching. When the tentative list is complete, some answers will be found to have a value of 3, others of 2, others of 1, and still others, that have no merit, of zero. It is inadvisable to assign minus scores for answers that are poorer than those which merely have no positive merit (although such answers will be found on occasion), as a minus score has been found to arouse a feeling of antagonism on the part of pupils which had best be avoided. It is wise to give a score of zero to all answers of either no merit or of a distinctly negative merit.

In the list of answers given above to the first statement the rearrangement in order and degree of merit was as follows (the number indicates the amount of credit allowed by the teacher) :

3 Able to stand heat and know the language.

2 Know language.

1 Stand heat.

1 Weather.

0 They gather rubber.

0 They live in Pará.

0 They live in Jungle.

0 Can stand smoke of palm leaves and nuts.

0 They are not English.

0 They are English.

With the scoring sheet before him for the first statement, the scorer is ready to score the first statement throughout all the papers, giving the various score values according to the answers on the sheets, and by constant reference to the tentative score sheet the same values can be given to all answers of equal merit. The scorer should watch carefully for new types of answers, and when such are found he should note them on the tentative sheet, with their assigned values, so that if similar answers recur later the same values can be given. The subjective judgment of the scorer determines the varying merit of answers; but once that value is determined, the valuation is kept constant. There might be some advantage in having the pupils themselves evaluate the varying answers; this can be done under conditions of pupil-scoring, but with teacher-scoring the delay necessary would usually be likely to overbalance the advantages. If there is unfairness in the subjective judgment, the unfairness weighs no more heavily on one paper than on another, and by these means a constant justice is obtained. Moreover, when all the papers have been scored and the relative difficulty of the questions has been determined, as is described in a later chapter, the teacher will have a means of determining where flagrant injustices have occurred and will then have the opportunity to make such corrections as will seem advisable.

When the scoring of the first statement has been completed, the second statement can be scored in similar fashion. The scorer makes a brief survey of the answers of the first half of the papers, arranges the answers on a tentative score sheet in their order of merit, arbitrarily assigns appropriate values for the various answers according to his best judgment of their merit, and finally scores the papers for the second question according to the table which he has made out.

When that is finished, the scorer repeats the operation with the remaining statements until all are completed. For the illustration which has been used here the completed score sheet, from which all the papers were scored, is given below.

SAMPLE SCORE SHEET FOR JUDGMENT TEST GIVEN ABOVE

Statement 1

- 3 Able to stand heat and know the language
- 2 Know language
- 1 Stand heat
- 1 Weather
- 0 They gather rubber
- 0 They live in Pará
- 0 They live in Jungle
- 0 Can stand smoke of palm leaves and nuts
- 0 They are not English
- 0 They are English

Statement 2

- 3 Keep insects (poison insects — mosquitoes) from biting them
- 2 There are insects there
- 2 So flies cannot get in
- 0 So hot
- 0 *They* bite them

Statement 3

- 3 Trees are dry of sap — sap all gone
- 3 Trees have died
- 1 They have taken all of it
- 0 To get (gather) sap for us
- 0 Fresh sap makes fresh rubber
- 0 Trees are cut down

Statement 4

- 3 Runs through Jungle where there are wild animals
- 2 Presence of wild animals
- 2 Dense Jungle
- 1 Flies, insects, etc.
- 0 Swampy
- 0 Largest, deepest, widest, longest river
- 0 Bridges

Statement 5

- 3 Smoke from palm nuts
- 1 The fires are so strong
- 0 The sap is strong
- 0 Rubber gets smoky
- 0 Palm nuts so strong
- 0 Palm leaves so strong

Statement 6

- 3 Ground covered by water
- 3 Too swampy
- 2 Get wet
- 1 Long rains
- 0 Trees grow
- 0 It is winter time
- 0 Rain kills the trees
- 0 Coconuts fall on them

Statement 7

- 3 Hot and unhealthy
- 3 Dangerous
- 2 Too hot
- 2 Dangerous animals or snakes
- 0 They can't get baby tigers. (The result of a strong suggestion from local zoo campaign)

The final score for any paper consists of the sum of all the values received on the individual questions. These are added up, and the final score is put in the place assigned. As has been suggested, if the score is placed in one corner of the sheet it can be hidden by turning down that corner and making a crease in the paper. The sheet is then ready to hand back to the pupil who wrote it and whose name should appear on the back.

Although, superficially, it might be thought that success in this test depends in great measure upon the pupils' abilities in other fields, such as literacy, legibility, spelling, penmanship, and the like, this need not be the case save within narrow limits. The field to be tested here is primarily that of geography, and the teacher should bear in mind that

handwriting, penmanship, spelling, grammar, and the like belong in other fields. This should not, of course, condone poor work in these fields, and the teacher can use as proofs of need there any suggestions which he may get, but the score in geography should indicate knowledge of geography and not of spelling, writing, and the like. As illustrative of this the following three answer papers for this test are given. The first paper is that written by the pupil who gained the highest score on this test. It was well written in all respects, except in one or two instances where the stress of test conditions probably induced the omission of a word or two. The second paper is that written by a foreign child whose knowledge of spoken English exceeded his ability in written English. It will be noted that this paper also received a high score. The difference between these two papers in the points mentioned above should be obvious. The second paper illustrates the case of a pupil unusually deficient for his grade in the mechanics of writing, spelling, and the like, but, for his class, unusually competent in the value and clarity of his ideas. His difficulties do not lie in geography but in orthography. In the third illustration is shown a pupil's paper which received a relatively low score. The handwriting and spelling were generally excellent, but the quality of the answers should be compared with the quality of those in the two previous illustrations. In all three illustrations the papers are reproduced as accurately and completely as print and the legibility of the original papers will permit.

SAMPLE ILLUSTRATION OF ANSWER SHEET, PUPIL 1¹

- (3) 1. Because They can stand the heat. and understand the language.
- (3) 2. Because so they do not get bite by the bugs.
- (3) 3. Because the Indians get all the sap out of the trees.

¹ This paper received the highest score in the class; it was well written and well spelled. The values given to the individual answers are shown in parentheses before each statement. The total value assigned the paper was 21.

- (3) 4. Because it runs through the jungle and are animals on trees waiting to kill people.
- (3) 5. Because the smoke to dry the rubber is so strong.
- (3) 6. Because they can't have any shelter because they have only canoes. and the ground is wet.
- (3) 7. Because they are afraid of the animals and cannot speak their lanuage and cannot stand the heat.

SAMPLE ILLUSTRATION BY PUPIL 2¹

- (3) 1. becus tha can sand the hot. and tock the langreg (Because they can stand the hot weather (or heat) and talk the language.)
- (3) 2. becus the mosquto are posen. (Because the mosquitoes are poisonous.)
- (3) 3. becus the tree is drie
- (2) 4. becus thear ar jaqust and monkisy and all Kind of amantes jump on pepol (Because there are jaguars (?) and monkeys and all kinds of animals (to) jump on people.)
- (3) 5. becus the smook is so strong. (Because the smoke is so strong.)
- (3) 6. becus it is to sompe to take the sap (Because it is too swampy to take the sap.)
- (2) 7. becus it is to hot and tha can tock thear langreg. (Because it is too hot and they can't (?) talk their language.)

SAMPLE ILLUSTRATION OF ANSWER SHEET BY PUPIL 3²

- (0) 1. because they get rubber for us
- (2) 2. so the fly wooden bite they
- (0) 3. because they ~~the~~ so they get rubber for us
- (2) 4. because lion tigar aninal are near the river
- (0) 5. because the palm nut blear then
- (0) 6. because cane nut fell Down on they head and kill then
- (3) 7. because is danaorgs and aninal are in the Jungle

¹ This paper received a score of 19. The writing was clear and distinct, but the spelling was as shown. The answers are rewritten, in parentheses, to make the meanings clearer.

² This paper was legibly written, and each word was quite plain. Margins were carefully observed, and the paper showed many signs of neatness and care. The total score was 7.

Values of the Judgment Test. As in the case of the True-False Test, one of the great values to be gained from the testing by this method lies in the objective review which is made possible when the papers are handed back to the pupils. Each statement should be carefully repeated, if either of the first two methods has been used, and a general discussion of the quality of the various answers should follow. Out of this discussion should come a rather clear idea of what the test is designed to bring out, and the pupils should receive and carry away with them the essential truths.

From another point of view the test offers to teachers a better way of gaining the good results which come from the traditional informal examination, since the test measures in a more exact way the ability of pupils to express themselves and does not encourage bluffing or verbosity or place too great stress upon orthographic correctness.

In addition, as will be shown in a later chapter, the test offers to the teacher a means of evaluating his teaching, and it provides a basis for making changes in the methods or the materials which are being used.

Chapter summary. The Judgment Test measures in a somewhat better way than others described the ability of pupils to recognize the principles upon which certain facts are based. It involves not only the ability to recognize the fact and the principle but also the connection which exists between them. It is constructed by making a series of true statements and asking for the reasons which make the statements true.

The test can be given in any of the three ways described in connection with the True-False Test, but is marked in a different way, giving a range of values to answers of varying merit. It is more largely subjective than the True-False Test in this matter, but a correction is made in a definite effort to distribute the effects of that subjective judgment equally on all pupils, thereby making it as fair or as unfair for one as another.

The test can be made objective with respect to the particular phase of the subject which is being used, and errors on the part of pupils in the mechanics of writing and spelling can be disregarded as far as the pupils' scores are concerned.

Sample Judgment Papers

SEVENTH-GRADE HISTORY PAPER¹

RAW SCORE_____ DATE_____ M SCORE_____

This is a Judgment paper. In the blank lines below write a short sentence which will tell the best reason you know why each of the statements is true. Do your best. Answer every statement you can. Write clearly.

1. The French and Indian War taught the colonists the importance of united action.

Answer _____

2. Underlying the political causes of the Revolution was a fundamental economic cause, the colonial system.

Answer _____

3. Samuel Adams won the title of "Father of the American Revolution."

Answer _____

4. The Townshend Acts were as dangerous to liberty as the Stamp Act itself.

Answer _____

5. Trade between Great Britain and her colonies fell off rapidly as a result of the colonial policy of nonintercourse following the Stamp Act.

Answer _____

¹ This test was constructed and used by Miss M. Beatrice Louy, McKinley School, Toledo.

6. King George III claimed that the colonists were represented in Parliament, even though they did not vote for its members.

Answer -----

7. After the "Boston Tea Party" the English government soon realized that it could not single out Boston for punishment.

Answer -----

EIGHTH-GRADE HISTORY PAPER ¹

RAW SCORE ----- DATE ----- M SCORE -----

This is a Judgment paper. In the blank lines below write a short sentence which will tell the best reason you know why each of the statements is true. Do your best. Answer every statement you can. Write clearly.

1. Our government has regarded unrestricted immigration as a hindrance to the welfare of its people.

Answer -----

2. Arbitration is a method often employed by nations in settling their disputes.

Answer -----

3. Nations have often acquired territory by purchase.

Answer -----

4. The South considered the Reconstruction policy dangerous to the best good of its white population.

Answer -----

5. A nation justifies the laying waste of an enemy's land as a war necessity.

Answer -----

¹ This test was made and used by Miss Isabel M. Smith, Harvard School, Toledo.

6. Americans are realizing that business methods should prevail in the Federal, state, county, and city governments.

Answer -----

7. Our national Congress has felt that infant industries should be protected.

Answer -----

FOURTH-GRADE HEALTH HABITS ¹

RAW SCORE ----- DATE ----- M SCORE -----

This is a Judgment paper. In the blank lines below write a short sentence which will tell the best reason you know why each of the statements is true. Do your best. Write clearly.

1. Plenty of sleep is necessary for good health.

Answer -----

2. A child needs more sleep than a grown-up person.

Answer -----

3. Going to bed at a regular time is an important health habit.

Answer -----

4. Everyone should sleep with windows open.

Answer -----

5. Children should not sleep on high pillows.

Answer -----

6. It is best not to play hard or exciting games just before going to bed.

Answer -----

7. People who eat hearty suppers late at night or drink tea, coffee, or cocoa just before going to bed may not sleep well.

Answer -----

¹ This test was made and used by Miss Rose Clippinger, Jefferson School, Toledo.

FIFTH-GRADE GEOGRAPHY PAPER ¹

RAW SCORE----- DATE----- M SCORE-----

This is a Judgment paper. In the blank lines below write a short sentence which will tell the best reason you know why each of the statements is true. Do your best. Answer every statement you can. Write clearly.

1. If a Kansas farmer should go to Argentina, he would know more about how to make a living than if he should go to northern Brazil.

Answer -----

2. It is very important that the United States cultivate trade relations with South America.

Answer -----

3. Hard rains in northern Chile would cause most of the fertilizer factories in the world to shut down.

Answer -----

4. The largest cities in South America are located on the Atlantic coast.

Answer -----

5. Brazil is sometimes called the young giant of the Western Hemisphere.

Answer -----

6. There are very few cities of importance in the interior of South America.

Answer -----

7. There is but one railroad in South America which connects the eastern and the western coast.

Answer -----

¹ This test was made and used by Miss June Mapes, special teacher, Board of Education, Toledo.

SIXTH-GRADE GEOGRAPHY PAPER ¹

RAW SCORE _____ DATE _____ M SCORE _____

This is a Judgment paper. In the blank lines below write a short sentence which will tell the best reason you know why each of the statements is true. Do your best. Answer every statement you can. Write clearly.

1. The leading industry of the Pacific states is agriculture.

Answer _____

2. Pennsylvania is the center of the coal and iron industry of the United States.

Answer _____

3. Cotton manufacture is one of the most important industries of New England.

Answer _____

4. Many crops are now raised on formerly arid lands in the plateau states.

Answer _____

5. Nearly one half the population of the United States is in cities or villages of more than 2500 inhabitants.

Answer _____

6. Chicago is the greatest meat-packing center of the United States.

Answer _____

7. Texas is the leading state of the Union in cotton production.

Answer _____

¹ This test was made and used by Miss Edna L. Roemer, Auburndale School, Toledo.

CHAPTER V

THE SELECTION TEST

Uses of the Selection Test. It frequently happens that a teacher feels the need of a quick means of measuring the progress of his pupils. The overcrowded curriculum, with its many exactions upon both the pupils and the teacher, means that it is often difficult for the teacher to find time not only for the teaching, which must be done, but also for the checking up and inquiry as to what has been actually taught, which should be done. The result is that both teacher and pupils may proceed for some time without any accurate knowledge on the part of the teacher as to the real achievement of many of the pupils and without much appreciation on the part of the pupils as to their own success or failure. Such a condition is undoubtedly as difficult as it is frequent; and if teachers could find a short and quick way to measure periodically the achievement of their pupils, it would be found to be mutually advantageous. The teacher could teach with a surer hand and with greater confidence, and the pupils would have a measure of their own efforts and an indication of their own progress.

The Selection Test, here described, offers to teachers such an opportunity. It is a test which may take various forms and is capable of much variation by the teacher himself to suit his individual needs or those of his classes. The chief element necessary for success is for the teacher to have a sure knowledge of the method of constructing the test, and an appreciation of the dangers that are present in testing in general and in these tests in particular. The test is easy to construct. A few minutes of spare time or a few notes made after a class has recited may form the basis for the test. A

few additional minutes in construction, following the procedure here advised, makes the test ready for administering. It is a test that children enjoy taking. Many of them find in it a type of game or puzzle which they take particular pleasure and satisfaction in solving. As soon as children find that the chief necessity for the proper solution of the test is a sure knowledge of the elements which have formed their study, an added incentive for that study (which is not at all undesirable) is provided. In most cases where the test has been used the pupils have attacked it eagerly with an attitude that is most favorable for learning and for measuring.

The test may be made difficult or easy, according to the circumstances which bring it about, and the teacher will find it readily adaptable to his needs. It may be as short as time may dictate, or as long as expedience may permit. It may form a part of a general testing program, such as is suggested in a later chapter, or it may, with equal ease, be an isolated unit for the purpose of acting as a guidepost for the teacher and the pupils. It is adaptable to many types of subject matter and covers a wide range of school activities. It is extremely easy to administer, and for most purposes it is equally easy to score. It forms a splendid basis for quick reviews, and, as will be shown, tests an important phase of school achievement which is difficult to measure by other means, namely, the ability to organize and arrange a given series of facts.

Varieties of Selection Tests. There are many variations of the fundamental character of this test which the ingenious teacher will discover for himself. Most of these can be classified under one or more of the four types which have been isolated and are illustrated below. The various types are adapted for use with different forms of subject matter and form a battery of possibilities of wide extent.

The teacher may also find that it is possible to use these forms in making other types of tests, as is shown in the illustrative test on page 337.

TYPE I. TWO-COLUMN SELECTION

The first type consists of a two-column selection of a series of pairs of related facts. Its construction is as follows:

STEP 1. SELECTION OF SUBJECT MATTER

The first step in the construction of this type of test is the selection of a well-defined unit of subject matter. The unit in this illustration formed a part of a battery of tests used in sixth-grade geography relating to the Pacific States. This part of the test was concerned with the meanings connected with some of the larger centers of population of these states.

STEP 2. SELECTION OF SERIES OF RELATED FACTS

The second step consists of the selection of a series of related facts or statements. These may be conveniently placed in the form of simple sentences of much the same essential character as is shown below. In one phase of the sentences one general type of idea should be followed through, whereas in the second phase a second type of idea should prevail. For the unit of geography given above the following sentences were constructed:

1. Los Angeles is the metropolis of southern California.
2. San Francisco has one of the best harbors in America.
3. Spokane is an important smelting center.
4. Portland is both a seaport and a river port.
5. Seattle is the largest city of Washington.
6. San Diego is a famous health resort.
7. Tacoma is a wheat-milling center.
8. Astoria is a famous fish-canning center.
9. Bakersfield is in the center of the oil district of California.
10. Sacramento is one of the oldest cities of this section.

In the list of sentences given above are found the qualities necessary for a test of this type. There are ten simple sentences. One phase of each of these sentences is of a single

type, in this case the subjects, which are invariably the names of cities of the Pacific states. Another phase of the sentences is also essentially a single type, the predicates, which tell what the various cities are noted for. They are, in fact, a series of related facts.

STEP 3. SEPARATION INTO TWO SERIES

The next step is to separate these pairs of related facts into two columns or series, which is done by merely disjoining the sentences at the verbs, with the elimination of the verbs themselves. The two columns will then appear as follows:

COLUMN I	COLUMN II
Los Angeles	The metropolis of southern California
San Francisco	One of the best harbors in America
Spokane	An important smelting center
Portland	Both a seaport and a river port
Seattle	The largest city of Washington
San Diego	A famous health resort
Tacoma	A wheat-milling center
Astoria	A famous fish-canning center
Bakersfield	Center of the oil district of California
Sacramento	One of the oldest cities in this section

STEP 4. ARRANGEMENT OF SERIES IN CHANCE ORDER

The fourth step is to take the items listed in the second column and rearrange them in chance order, giving them, at the same time, letters as distinguishing marks. The items in the first column may be left as they are and numbered. The arrangement in chance order may be conveniently done as in the previous tests. The numbers of the statements, or the statements themselves, can be placed on cards or small slips of paper, the cards thoroughly shuffled and blindly drawn one after another. As they are drawn, they are placed in turn in the second column. Arranged in this way, the first column left in its original order and numbered, the

second column rearranged in chance order and lettered, and the whole arranged for administration by the mimeograph method, the statements developed become as follows:

Sample Selection Paper of Type I

SIXTH-GRADE GEOGRAPHY PAPER¹

RAW SCORE----- DATE----- M SCORE-----

This is a Selection paper. In the spaces in front of the numbers in Column I place the letters of the phrases in Column II which best explain them. Write the letters clearly in the spaces. Do your best.

COLUMN I

COLUMN II

- | | | |
|------|------------------|---|
| ---- | 1. Los Angeles | A. Famous fish-canning center |
| ---- | 2. San Francisco | B. A famous health resort |
| ---- | 3. Spokane | C. Metropolis of southern California |
| ---- | 4. Portland | D. One of the oldest cities in this section |
| ---- | 5. Seattle | E. A wheat-milling center |
| ---- | 6. San Diego | F. One of the best harbors in America |
| ---- | 7. Tacoma | G. Center of the oil district of California |
| ---- | 8. Astoria | H. Both a seaport and a river port |
| ---- | 9. Bakersfield | I. An important smelting center |
| ---- | 10. Sacramento | J. Largest city of Washington |

Be sure that your name is on the back of this sheet.

SPECIAL CONSIDERATIONS FOR TYPE I

The test is now completed and is ready for use. This consists essentially in having the pupils place in front of each of the numbers preceding the names in Column I the letters preceding the statements or phrases in Column II which

¹ This test was constructed and used by Miss Laura Kuhr, Newton School, Toledo, Ohio.

pair best with them. This is not so easy as it might seem on the surface, because there is a somewhat fine discrimination involved, which, if neglected, may lead some children astray. For instance, in the above test 7. *Tacoma* may with equal correctness be classed as *E. A wheat-milling center* or as *I. An important smelting center*. If a pupil, however, gives Tacoma as an important smelting center, he will find himself unable to find a suitable pair for *Spokane*. Again, both *San Diego* and *Sacramento* may fit well with *D. One of the oldest cities in this section*; but if that statement in Column II is paired with *San Diego*, the pupil will find a very lame pairing with *Sacramento* and *B. A famous health resort*. The teacher should try to anticipate these possibilities in the pairings and, if it is found necessary, to make such changes as are necessary to reduce them, unless, as in this case, the presence of such pairs helps to make the test more valuable. As a rule such variations in the possibilities of answering are desirable, unless the statements are so general that almost any statement may fit properly in the selection, in which case no real test is involved. If this should happen, the test should be made more specific. It may be added that, when scoring, the pupils should be given credit when their answers are right, even though the answers may not fit with the first group of paired statements as made out by the teacher. In the above case, for instance, if 7 were matched with either E or I, it should be counted as correct, but if matched with I, 3 would be wrong. So also with 6 and 10: if D were matched with either 6 or 10, it would be correct, but if it were matched with 6, then 10 would be forced to be incorrect. The method of scoring is similar to the general methods used with other types of this test and is described in a later section of this chapter.

This type of Selection Test has some special values. For work similar to that above, where it is wished to have a measure, quickly taken, of the range of exact knowledge which a group of pupils has acquired, the type of test just

given will be found of great value. If given in connection with the Judgment Test described in the preceding chapter, the greater values of the older types of school examining will be gained and at the same time most of the injustices of the older type of test will be avoided. The two tests given together provide an opportunity for self-expression, discrimination of a rather nice sort, and an exhibition of the degree of exactness of knowledge which the class possesses.

It is possible to adapt this test in many ways to local or special teaching conditions. In this connection attention is called to the special form given in Chapter XVII, p. 338. Here is shown a form of the test applied for use in the second grade, and a method is there given for eliminating certain of the more disturbing mechanical difficulties which the test, in the form shown above, presents to pupils in the lower grades.

TYPE II. REARRANGEMENT TEST

A further variation of the Selection Test may be classed as Type II. It is particularly useful when one wishes to test the ability of pupils to rearrange, reorganize, or identify a series of facts as a series, rather than to test the knowledge of the facts themselves in other relations, although that is, of course, to a degree, necessary for a correct rendering of the test. In the example given here a form of reorganization of a series of facts is illustrated. The teacher using it in the classroom can adapt this form in many ways, with different kinds of materials, and can vary the usage so as to make a great variety of forms.

STEP 1. SELECTION OF SUBJECT MATTER

The first step, as in the preceding type, is to select the subject matter that is involved. In the case cited below, the test given comprised part of a battery of tests in eighth-grade history of the United States. The effort was made to

test certain types of organization ability which the teacher had tried to develop, rather than to measure only a recognition of the facts themselves.

STEP 2. MANUFACTURE OF STATEMENTS

The next step is the selection, based upon this subject matter, of a series of statements or facts which taken together form a connected whole or else have some particular organized idea that is common to them. Because of the mechanics of scoring this test it has been found wise for a single test to contain five of these series of statements or facts, each of them to consist of four elements. The five series of events, arranged in chronological order within each series, as selected for this test, are given below. Each series is labeled for identification purposes.

FIVE SERIES OF STATEMENTS IN CHRONOLOGICAL ORDER

1. *Invention*

Invention of the cotton gin	1794
First trip of the <i>Clermont</i>	1807
First steam locomotive used on the Baltimore & Ohio Railroad	1830
Invention of the McCormick reaper . . .	1831

2. *Western expansion*

Completion of the Erie Canal	1825
Admission of Texas into the Union . . .	1845
Discovery of gold in California	1848
Completion of the first transcontinental railroad	1869

3. *Slavery*

Missouri Compromise	1820
Petitions for abolition introduced into Congress	1834
Dred Scott decision	1857
Emancipation Proclamation	1863

4. *Civil War*

First battle of Bull Run . .	July,	1861
The <i>Trent</i> affair	November,	1861
Gettysburg	July,	1863
Appomattox	April,	1865

5. *Reconstruction*

The Reconstruction Act . .	March,	1867
President Johnson acquitted of impeachment	May,	1868
Acceptance of Fourteenth Amendment	July,	1868
Enfranchisement of negroes	March,	1870

STEP 3. DISARRANGEMENT OF STATEMENTS WITHIN SERIES

The third step is to disarrange, within each series of events, the chronological or other order which forms the basis of the test. This may be done, as in the previous tests, by placing the numbers on slips of paper or cards, shuffling the cards, and drawing them in rotation. As the numbers are drawn, they may indicate the order in which the statements are finally to be placed. In this case the final form in which this test was drawn and assembled is shown in the form given below. The test is presented in the form in which it would be arranged to be given by the mimeograph method. The general arrangement of the form may be noted as similar to that previously used, although other forms might be devised, such as including an entirely separate space for copying and regrouping the statements. Any form which gives a certain space for a short and easily read answer is better, however, than a form which has to be carefully read in correction, especially when that reading adds nothing save a rearrangement of statements. In the form as shown below nothing need be read in correction except the rotation of the numbers in the outlined spaces as they have been filled in by the pupils according to the directions.

Mimeographed Form of Rearrangement Test. Type II

EIGHTH-GRADE HISTORY PAPER ¹

RAW SCORE_____ DATE_____ M SCORE_____

This is a Selection paper. In each group write 1 for the earliest event or statement, 2 for the next one, 3 for the next, and 4 for the most recent. Be careful to do your best.

Group I

- The first trip of the *Clermont*.
- Invention of the McCormick reaper.
- First steam locomotive used on the Baltimore & Ohio Railroad.
- Invention of the cotton gin.

Group II

- Discovery of gold in California.
- Completion of the Erie Canal.
- Completion of the first transcontinental railroad.
- Admission of Texas into the Union.

Group III

- Dred Scott decision.
- Petitions for abolition introduced into Congress.
- Emancipation Proclamation.
- Missouri Compromise.

Group IV

- First battle of Bull Run.
- The *Trent* affair.
- Appomattox.
- Gettysburg.

Group V

- Acceptance of the Fourteenth Amendment.
- President Johnson acquitted of impeachment.
- The Reconstruction Act.
- Enfranchisement of negroes.

Be sure that your name is on the back of each sheet.

¹ This test, prior to certain modifications, was constructed and used by Miss Nettie Fehn, Newton School, Toledo.

SPECIAL SCORING DIRECTIONS FOR TYPE II

This type of the Selection Test requires a special form of scoring, because the regular form of scoring these tests, as outlined in a later section of this chapter, is unfair. The injustice may be illustrated in the following example.

In the test as given above let it be supposed that a pupil does the first series in the test correctly in the following manner. Only numbers are given to prevent confusion.

CORRECT PLACING	PUPIL'S PLACING	ELEMENTS OF SERIES I
2	2	1
4	4	2
3	3	3
1	1	4

Let it now be supposed that a certain positive credit is allowed for each element that is correctly placed, as is recommended in general in the later section of this chapter above referred to. In this case, comparing the first two columns in the example given above, it is easy to see that the pupil has put in the proper position each of the elements and that he should have, let it be assumed, a total score of 4 for the series.

Now let it be assumed that a second pupil has placed the same elements of Series I in the following manner:

CORRECT PLACING	PUPIL'S PLACING	ELEMENTS OF SERIES I
2	1	1
4	3	2
3	2	3
1	4	4

Here, in comparing the first two columns it can be easily seen that the pupil has not placed any single one of the elements in its rightful position, and according to the scheme of scoring as outlined for the preceding example the score for the series should be zero. A closer examination of the placing by the pupil will, however, reveal this fact, that in

reality only *one* and not *four* of the elements have been wrongly placed. It is clear that the pupil knew nothing about the relative time in which the cotton gin was invented, as he has placed that last, but it is also clear that he has placed the other three elements in rightful position with each other.

If the reader cares to work out on paper the varying combinations of four numbers, 1, 2, 3, and 4, with the four correct placings as given in the two examples above, he will find that instead of *absolute wrongness* it is possible to have *relative wrongness* (or rightness, according to the attitude taken) and that this relative wrongness is in proportion as one (any one), any two, any three, or the entire four elements are actually, and not as in the above case superficially, wrongly placed. In addition, the degree of misplacement should be taken into consideration. From this point of view, and in harmony with the ideas of fairness as expressed throughout this book, it is unfair to give these varying degrees of wrongness the same unvarying score, namely zero. A method is given below which gives credit where credit is due, not in terms of absolute rightness and absolute wrongness but rather in terms of what is actually done, or partially done, correctly.

The degree of rightness or of wrongness in the placing of a group of elements such as those in this test is best represented in the amount of difference there may be between the pupil's placing of each element and the correct placing. Even if, for example, only one element were wrongly placed, it should make a difference whether it is one or three places away from its correct position. The method is as follows:

A key should be made, to be placed beside the series to be corrected. In the examples cited above, this key is represented in the column headed "Correct Placing." Then the scorer should make mental (or, if needs be, actual) note of the differences of each of the pupil's placings from that of the key, and finally add these differences together for a final

sum of the differences of the pupil's placings from that of the correct placing. It will be noted that this takes into account not only the fact of wrongness but also the degree of wrongness. Then, with the sum of these differences in mind the scorer can turn these differences (which are measures of wrongness) into measures of rightness by consulting the following table. This table is short and can be easily remembered by the scorer after a very few series have been scored, especially if the above directions for making the test have been faithfully followed. The table cannot be successfully used with series of *more* or of *less* than *four* elements.

TABLE FOR CONVERTING SUMS OF DIFFERENCES INTO SCORES IN
REARRANGEMENT TEST

When the sum of differences is 0 the score is 4
When the sum of differences is 2 the score is 3
When the sum of differences is 4 the score is 2
When the sum of differences is 6 the score is 1
When the sum of differences is 8 the score is 0

In the following examples the method of using this table is shown. Let it be supposed that the key for a certain group or series in a Rearrangement Test and also the pupil's answers to that series are as given in the tables below:

KEY ANSWERS	PUPIL'S ANSWERS
2	3
4	4
1	2
3	1

In these answers the pupil has made the following errors and has shown the following differences from what should be rated as a correct answer :

The first statement or event should have been placed second. The pupil has placed it third, a difference of one place on this item. The second statement or event has been placed fourth by the pupil, which is its correct placing. The

third statement or event has been placed second by the pupil whereas it should have been placed first, a difference of one place. The fourth statement has been placed first by the pupil and should have been placed third. The difference here is two places. The scorer would determine these differences and would add them together as follows :

KEY ANSWERS	PUPIL'S ANSWERS	DIFFERENCES
2	3	1
4	4	0
1	2	1
3	1	2

If the numbers in the "Differences" column in the preceding table are added together, the sum of differences necessary to find the final score for the series is secured. In this case the sum of the differences is 4, which, by consultation of the table given on page 101, is found to be a score of 2.

In the following case the problem is worked out in similar fashion :

KEY ANSWERS	PUPIL'S ANSWERS	DIFFERENCES
3	3	0
2	1	1
1	2	1
4	4	0

Sum of differences, 2 ; score, 3. (See table, p. 101.)

TYPE III. REGROUPING TEST

A third type of Selection Test which has certain values for the classroom teacher is that which is here called a Regrouping Test. In contrast to the ability which is necessary to cope successfully with the first and fourth types as they are here given, that of selection or matching of related facts, or of the second type (which has just been outlined), this type is primarily concerned with the identification and regrouping of a series of facts or of groups of facts having a common factor.

STEP 1. THE SELECTION OF THE FACTS

The first step consists essentially in the selection of the groups of facts of which the test is to be made up, and is followed by a specific selection from each group of a number of facts suitable for the test itself. In the illustration given below the test formed part of a battery of tests in history, the other portion of the battery being used as illustrations in describing other forms of tests. The object was to test the ability of the pupils to identify and group representative war leaders in this country. The groups selected were leaders in the Revolution, the War of 1812, the Civil War, the Spanish-American War, and the World War. The leaders selected were ten in number and are given in the following list :

- | | |
|----------------------|-----------------------------|
| 1. Israel Putnam | 6. U. S. Grant |
| 2. Anthony Wayne | 7. George Dewey |
| 3. Isaac Hull | 8. Winfield S. Schley |
| 4. Oliver H. Perry | 9. Theodore Roosevelt (1st) |
| 5. David G. Farragut | 10. John J. Pershing |

To increase the selectiveness necessary to complete the test and thereby to increase the difficulty of the test for school pupils in the upper grades, it is sometimes desirable to add to the matters legitimately a part of the test certain elements otherwise unconnected with it. In the example here given it was decided to add to the foregoing list of war leaders half that number, leaders with whom the pupils were familiar but who were not of the United States. This list of leaders was as follows :

1. Cornwallis
2. General Burgoyne
3. Santa Anna
4. Douglas Haig
5. Hindenburg

STEP 2. FORMATION OF ITEMS IN CHANCE ORDER

With the selection of the names of the war leaders as given above the test is ready for assembly, which consists merely in the placing of the items in chance order. In this case chance order is achieved in the way that has previously been described. The names, or their numbers, may be written on slips of paper, the slips shuffled, and the names of the leaders placed in the order in which the slips may be drawn. One chance order is shown in the following illustration, which has been arranged for use with the mimeograph method.

Sample Selection Paper. Regrouping Test. Type III

EIGHTH-GRADE HISTORY PAPER¹

RAW SCORE----- DATE----- M SCORE-----

This is a Selection paper. The following is a list of leaders prominent in war. Select those who took part for the United States (mark "U. S." in Column I); then tell in which war they took part (write out name of war in Column II). Leave blanks after names of men who did not fight for the United States. Be careful to do your best.

	COLUMN I	COLUMN II
1. Cornwallis	-----	-----
2. Winfield S. Schley	-----	-----
3. U. S. Grant	-----	-----
4. David G. Farragut	-----	-----
5. Santa Anna	-----	-----
6. General Burgoyne	-----	-----
7. John J. Pershing	-----	-----
8. Isaac Hull	-----	-----

¹ This test was constructed and used as part of a battery of tests by Miss Nettie Fehn, Newton School, Toledo.

9. George Dewey	-----	-----
10. Douglas Haig	-----	-----
11. Theodore Roosevelt	-----	-----
12. Oliver H. Perry	-----	-----
13. Israel Putnam	-----	-----
14. Hindenburg	-----	-----
15. Anthony Wayne	-----	-----

Be sure that your name is on the back of this sheet.

SPECIAL SCORING DIRECTIONS FOR TYPE III

In scoring this paper the general directions as formulated later in this chapter can be followed, and in addition to giving a single score for each element a score should be given for each part of the answer which is correct. Thus, if 8. *Isaac Hull* is recognized as an American leader in war, but is not recognized as taking part in the War of 1812, one part only should be given credit. If, on the other hand, both parts are correct, credit should be allowed for both. It is probably unwise to give a minus score for leaders who are attributed to the United States but who were not leaders for the United States, or for mistakes in identifying the particular war in which any particular leader took part. It is always better, if possible, to give a positive score for what is right than a negative score for what is wrong, if only because of the better attitude which accompanies the giving.

TYPE IV. SELECTION OF RELATED FROM UNRELATED FACTS

A further type of the Selection Test will be found useful not only for testing knowledge but also for offering variety to the testing program. Type IV offers a number of related and unrelated facts, with the problem for the pupil of select-

ing the related pairs. It is somewhat similar to Type I, but exhibits a different method; and where that test measures an ability to match related pairs among many possibilities, this test necessitates the selection of a suitable pair from a rather limited group of possibilities.

STEP 1. SELECTION OF SUBJECT MATTER

The first step in the construction of this type of Selection Test is similar to that of the other forms that have been described, and consists in the selection of the subject matter which is to be used as the basis for the test. In the illustration given below the subject matter was the geography of South America, and the specific purpose of the test was to determine the ability of the pupils to pick out of several possibilities the essential and relevant facts which were presented.

STEP 2. CONSTRUCTION OF PRELIMINARY TEST SENTENCES

The second step was to construct a series of simple and true statements with respect to the geography of South America which these pupils had studied. The teacher tried to have each statement include an idea involving something more than superficial knowledge. These statements, in the order of their construction, were as given in the list below:

1. The largest plateau of the Andes is called the Plateau of Bolivia.
2. The mouth of the Amazon is wide.
3. A llano is a plain.
4. Dense tropical forests grow where it is humid.
5. A llama is an animal.
6. Mestizos are mixed races.
7. Brazil grows two thirds of the world's coffee.
8. The city of Pará exports rubber.
9. One of the greatest industries of Argentina is grazing.
10. The most important product of Chile is nitrates.

STEP 3. CONSTRUCTION OF UNRELATED FACTS

The third step was to construct for each of the statements given above a series of three similar but incorrect and unrelated predicates. These, when completed, were as in the series given below, where the true predicate (or portion of the predicate necessary for the test) is found in parentheses.

1. (Bolivia) Chile — Peru — Brazil
2. (wide) shallow — small — narrow
3. (plain) tree — bush — animal
4. (humid) dry — cold — mountainous
5. (animal) fish — grassland — cloth
6. (mixed races) Spaniards — Portuguese — Indians
7. (two thirds) one quarter — one half — nine tenths
8. (rubber) shoes — coffee — cotton
9. (grazing) gold-mining — rubber culture — cotton-growing
10. (nitrates) gold — cattle — wheat

It will be noted that in each of the groups given above there are three words, or groups of words, which make, on first inspection at least, possible substitutes for the words of similar nature in each of the predicates of the preceding sentences. However, no one of the words in each group is as good or as true as the similar words in the original sentence.

STEP 4. CHECKING THE TEST

The next step is to check each of these last selected possibilities against the key word in the original sentence. If in any case it is found that the selection made offers as good a completion of the original sentence as the original sentence offers, there are two possible changes that may be made. In the first place, it is possible to change the original sentence to one which makes it either easier or more practicable to find a number of similar but unrelated words. In the second place, the selection of words may be remade so as to make the unrelated words really and not merely apparently unrelated.

STEP 5. CHANCE-ORDER DISTRIBUTION OF BOTH THE SENTENCES AND THE SELECTION UNITS

The fifth step consists of two things: the first, the recasting of the original sentences in chance order, so as to avoid any possible sequence in the statements; and the second, the attachment of the unrelated words, selected as in the third step, to the original sentences. When this has been done, this test now appears in its final form as follows. It has been arranged for use by the mimeograph method.

Sample Selection Test of Type IV

SEVENTH-GRADE GEOGRAPHY PAPER

RAW SCORE_____ DATE_____ M SCORE_____

This is a Selection paper. Put a line under the ONE out of the four possible endings in each sentence below, which you think is most true and makes the best sense. Do your best.

1. The city of Pará exports (shoes — coffee — cotton — rubber).
2. Dense tropical forests grow where it is (dry — cold — mountainous — humid).
3. A llama is a (fish — animal — grassland — cloth).
4. The mouth of the Amazon is (shallow — wide — small — narrow).
5. The most important product of Chile is (nitrates — gold — cattle — wheat).
6. A llano is a (tree — bush — animal — plain).
7. Brazil grows the following fraction of the world's coffee (one half — one quarter — two thirds — nine tenths).
8. The largest plateau of the Andes is called the Plateau of (Chile — Bolivia — Peru — Brazil).
9. One of the greatest industries of Argentina is (gold-mining — rubber culture — grazing — cotton-growing).
10. Mestizos are (Spaniards — mixed races — Portuguese — Indians).

Be sure that your name is on the back of this sheet.

It will be noted in this final construction that in addition to placing the sentences in order by chance the four possibilities at the end of each sentence have also been placed in order by chance. In this case a satisfactory chance order is achieved by merely not following any plan in the setting down of the four possibilities. It is not wise to place the right possibility deliberately first in the first sentence, second in the second sentence, and so on, or last in the first group, next to last in the second group, and so on, since such a set plan might be easily detected by the pupils and the value of the test would thereby be lost. The absence of any set plan makes the right answer on the part of the pupils either a matter of lucky guesswork, one chance in four, or else the result of deliberate and reasonable selection, which is, of course, the result which is desired. One way of making this chance order is to number four cards (perhaps the first four numbers which have been made and used on the cards for the True-False Test or one of the other tests) and to draw them one at a time after they have been thoroughly shuffled. An easy plan is merely to draw one card after each shuffling. Thus, if 3 is drawn, the right element of the selection would be placed third in the list of four; and if 1 were drawn, it would be placed first.

It is possible for other variations of this form of Selection Test to be used. One possible variation, for instance, might be to change the wording of the sentences to read as follows:

1. (Shoes — coffee — cotton — rubber) are (is) exported from the city of Pará.

2. The Plateau of (Chile — Bolivia — Peru — Brazil) is the largest plateau of the Andes.

Similar changes to these could be made with most of the other sentences given above, offering a convenient variation to use.

How to give Selection Tests. In addition to the special methods of giving Selection Tests which have been noted in

the cases of the foregoing illustrations of the various types, there are certain general methods of administration which may be found helpful to teachers who are using the tests. In the methods given below the teacher will find, perhaps, many that are not adapted to his needs. Teachers should therefore remember that these are suggestions, and not directions which cannot be changed. So long as the general needs are filled, and so long as the teacher is consistent in what he *does* do, he will find his own variations of these directions probably better suited to his own conditions, and he should not hesitate about making such adaptations as may seem wise.

Giving the tests by dictation. It is not possible in a test of this kind to use a dictation method such as has been previously described for use with other forms of Teacher's Classroom Tests, because much of the value of this kind of test lies in the sober reflective judgment and weighing of possibilities (together with the weighing of the various processes of trying out the possibilities) which are called for. When the teacher desires, then, to use the dictation method, it must be such a method as will allow all the pupils taking the test to write out practically the entire test before attempting to complete it.

As an introduction to dictating a test, only such directions should be given as may be well remembered by the pupils. For Type I (p. 93) this procedure might be as follows:

Please take out a sharp pencil (or pen) and a piece of paper, in order to be ready to write as I dictate. This is a paper which will tell me how well you know some of the things we have been studying about the geography of the Pacific states. I shall dictate two series of statements. One series consists of the names of cities of the Pacific states. This I shall call Column I. The other series consists of statements about these cities, and I will call it Column II. These are all mixed up, and I

shall want you to straighten them out ; so leave a wide margin when I dictate Column I to you, and you will find out later what to do with it.

Now write your name at the top of your paper, and after you have finished that write the date of today [pause]. When you have finished, turn over your paper so that your name will be on the back where you cannot see it [pause]. Now write the words "Column I," like this, in the middle of the line at the top of the paper.

At this point it would be wise for the teacher to illustrate on the blackboard just how to write down these numbers and names for Column I. After the pupils are familiar with the test, the directions can, of course, be much shortened, for the pupils will know what to do without being told specifically. When the test is new, however, the teacher can take nothing for granted and must be very careful to have the pupils follow the preliminary directions faithfully. It is also a good plan to have all the pupils use paper of the same kind, since it will save much time later in scoring.

Now I will dictate the names in Column I. When you write them, put down the numbers too and remember to leave a wide margin on the left.

This should be followed with an illustration on the blackboard:

---- 1. *Los Angeles*

The teacher should then dictate :

Number One : "Los Angeles, one."

Number Two : "San Francisco, two." Etc.

The dictation may be continued in this fashion until all the statements or names of Column I have been dictated. When that has been completed, Column II may be dictated, as follows :

I shall now dictate Column II. These have letters instead of numbers. Be careful to copy them just as I

read them to you. Write "Column II" under the last statement in Column I, in the middle of the line [illustrating].

"A. A famous fish-canning center, A.

"B. A famous health resort, B." Etc.

When the dictation of Column II has been completed, the pupils are ready to take the test and to receive the proper directions for doing so. These directions may be somewhat as follows:

You are now ready to finish this paper, and I want to see how well you can do. Find the name of the city that is a famous fish-canning center and put the letter "A" in front of it in the space you left in the margin. Then find the name of the city that is a famous health resort and put the letter "B" in front of it where you left the space. Are you ready? Start.

Type II can be dictated in much the same manner as has been given above for Type I. Because of the variety of ideas that are involved and also because of the variations in chronology between the groups, it is better, when dictation is used, to dictate only one group at a time and to allow the completion of any single group before dictating another.

Dictation of the Regrouping Test (Type III) is quite simple, since all that is required is the preliminary directions (which may be written upon the blackboard in order that they may be remembered), to be followed by the dictation of each element separately, allowing time after the dictation of each element for the rendering of an answer.

Type IV can be treated much in the nature of a True-False Test. In this case if the dictation is very slow, depending, of course, somewhat upon the abilities of the pupils, it would be possible for the pupils to make a satisfactory answer to the questions by writing down one word, the correct completion of the statement. This is because the selec-

tion lies within the sentence or statement itself and is not a selection from different statements. If the statements are read slowly enough to be retained by the pupils long enough, a wise selection may be made. There is also much to be gained from judicious repetition. The dictation of this test might be as follows, after the pupils had been asked to make ready for the test by producing the necessary materials:

I am going to read you a series of statements about the geography of South America which we have been studying. In each statement there are four words given from which you must make a choice to make the statement correct. Please write the one word or phrase that you consider to be correct, and do not write anything else except the number of that statement.

Number One: "The city of Pará exports shoes — or coffee — or cotton — or rubber." Which is correct? Don't tell me, but write it on your paper. Be sure to write down the same number that I read to you. I will repeat. Number One: "The city of Pará exports shoes — or coffee — or cotton — or rubber." Write.

Number Two: "Dense tropical forests grow where it is dry — or cold — or humid — or mountainous." Which of these is correct? [Repeat.] Write.

Advantages and limitations of the dictation method. The main advantage of the dictation method is that it takes very little time to prepare for the giving of the test, only so long as is necessary to prepare the statements. Thus, if a teacher wishes to give the test and, having prepared it, finds little time during school to make the necessary preparation for using one of the other methods that are outlined, it is possible to give the test to the pupils by dictation with no further preparation. The method, however, has two distinct disadvantages. The first of these is that by this method there is much chance for error on the part of the pupils in many ways. They may fail to hear the words clearly and

thereby mistake them for others. They may be unable to spell the words properly and when re-reading may recognize them as something different from what they really are. They may not make the numbering or the lettering correctly and thereby make it impossible to make the proper notation of their selections. In any event the chance for error aside from the ignorance of the pupils is so great that dictation should be used only when the results are not intended for any purposes save those of the pupils themselves. In the second place, dictation takes more school time than any of the other methods proposed and from that standpoint is much less efficient. Because of the fact that the entire test, in most cases, has to be copied from dictation, and that therefore the dictation must of necessity be slow, the amount of time which the pupils spend in unnecessary or unproductive labor is out of proportion to the time which they spend in their productive efforts to complete the tests. Wherever accurate and reliable results are desired, especially when the tests are to be scaled, as is shown in a later chapter, one of the other proposed methods will be found, as a rule, more useful.

Giving the test by the blackboard method. The second way of giving the test, that of writing it on the blackboard, is similar to that described in previous chapters. The writing should be done, if possible, when the pupils are out of the room, as at recess, before school opens in the morning, or when the pupils are in the gymnasium, unless the classroom teacher has to be with them at those times. It is also wise for the teacher to provide some sort of covering for the blackboard, such as has been described on page 41, to prevent the pupils from seeing the board and comparing notes before the test is ready to be given.

When the blackboard method is used, the tests can be written as has been explained and the directions can be given orally. In Type I the pupils need merely write down on their papers the numbers and insert the letters in the

proper places before the numbers after they have made their selections. In Type II the sequence of numbers is all that need be required, provided that each group of answers is properly labeled. In Type III the entire test can be copied or only the numbering need be copied, and the two columns can be filled in as in the form given. In Type IV the only writing necessary is the copying of the numbers and the writing of the selected words.

Criticisms of the blackboard method. The blackboard method is considerably superior to dictation in that it consumes far less of the class time for its completion and at the same time eliminates many of the sources of extraneous error. The main source of error, aside from the ignorance of the pupils, lies in the chance of misnumbering or mislettering the statements as they are taken from the board. It will be found, however, that the pupil who finds he has made such a mistake once will be very careful not to make it a second time. The blackboard method has the disadvantage of taking up much of the teacher's time. This is not a very serious consideration, however, because if the dictation method is used it takes as much of the teacher's time, if not more, to give the dictation as to write the same material on the blackboard. In addition, in the dictation method the time which the teacher spends may be multiplied by the number of pupils to find the actual amount of time which is unproductively spent. The time which the blackboard method actually saves the teacher is not appreciable either before or during the test, but in comparison with the traditional types of school examinations the time saved comes after the test is over, when the scoring begins.

Giving the test by the mimeograph method. In this test, as in all the other forms of tests that have been described, that method of giving it which brings the most reliable results, which is the easiest to score and handle, and which is fairest to the teacher and to the pupils as well is the mimeograph, or stencil, method. By this method every

pupil has before him on his desk a copy of the test with the complete directions for completing it and with the proper spacing and numbering. Each pupil, moreover, has exactly the same elements to start with and the same chances of having extraneous errors creep into his work. The result is likely to be more nearly a measure of his true ability in the test than if reached by either of the other methods.

In the mimeograph method the teacher must prepare very carefully the material from which the stencil or other master form is to be made, making very sure that there are included in the master form all the essential elements of the test. The checking should be carefully done. Each statement should be read and compared with the original. Each number and each letter designation should be checked against the original, and the spaces for the answers should be both adequate in size and correctly placed not only for answering but for speed in scoring. It is very much to the interest of the teacher to reduce to a minimum any possible sources of annoyance in after-correction and any possible difficulties due to typographical errors. After the whole test has been carefully checked for errors, and these errors corrected in so far as possible, it is ready to be given to the pupils. A number of examples of finished tests arranged for the mimeograph method are shown in this chapter.

If the test when completed has two or more sheets, it will be found wise to clip them together in groups. This saves the time of the teacher when the sheets are distributed, and prevents confusion during the working of the paper. There is a wide variation in the amount of time which different pupils require to complete these tests, and when the sheets are not clipped the confusion arising from the passing of pupils to the front of the room to hand in a completed sheet and to receive a fresh one should be avoided.

In giving the tests by this method there are two possibilities by the use of which the teacher can maintain a degree of impartiality in the scoring of the papers: One of

these has been suggested in the previous methods, where the pupils are asked to write their names on the backs of the sheets which they are working upon. By this method, unless a teacher is very familiar with even small samples of the pupils' handwriting, the individual pupils may remain anonymous so long as the teacher wishes. The second method is one which makes a chance turning over of the papers incapable of revealing the name of the writer of a paper. Here the teacher should prepare a number of cards sufficient for each member of the class to have one, and the cards should be numbered consecutively. These should be shuffled and passed out to the pupils, who then should write their names on the cards, and the numbers of the cards on their papers. The scoring can all be done by number, and the card may be referred to at the end for the return of the papers as well as for the identification of the pupils in need of special attention.

Scoring the Selection Test. In scoring the papers the teacher should be careful to refrain from any preconceived notions of what is or is not a correct answer to any element. In spite of careful checking and self-questioning when the test is constructed, it is easy for ambiguities to creep in, as well as for more than one series of correct answers to be possible. In the previous illustrations some of these cases have been pointed out, although in most cases the teacher will find that if there are two possible correct answers for some particular question, the selection of an answer different from that originally intended will probably mean that some other answer will of necessity be incorrect.

In this test, as in all tests, it is possible for the teacher to adopt a positive attitude toward correctness and add points to those papers of pupils who deserve them, rather than to adopt the traditional negative attitude, on the 100-percentile scale, which forces a teacher always to subtract points from the papers of undeserving pupils, or to the extent that pupils are undeserving. This positive attitude will be found

to make a great difference in the attitude of both the teacher and the pupils with respect to the giving and the taking of tests.

In Type I and Type IV the scoring is quite simple. A fixed credit of 2 or 3 points (preferably 2) can be assigned for each paired group or correct selection, the addition of all the separate credits so received constituting the final score of the pupil. A master key should be made, as has been suggested in other tests, which can be placed beside the answers of the pupils and which will make the scoring very much faster and more accurate than attempting to memorize the correct placings. All that may need to be remembered will be the credits to be given for answers which are possibly correct and which were not anticipated, as has been suggested in previous sections. For Type IV, especially in the mimeographed form, it will be possible for the teacher to make a key by taking a blank sheet of paper and cutting holes a little larger than the words which are correct and placing those holes in the same relative position as the correct placings. It will then be possible merely to superimpose the key sheet over the answers of the pupils and mark correct those answers which are underlined and which show through the holes in the key sheet.

Type III can be corrected in two ways, either by giving a positive score of, say, 2, for each element that is correct in both parts and giving no score for any element that is only half correct, or else giving a score for each part that is correct, regardless of the correctness of the corresponding part of the same element, as was described in the discussion of this type. As there stated, it is probably better to give a part of the credit to each element that is right, regardless of any other part.

Chapter summary. The Selection Test offers to teachers a medium for measuring the achievements of pupils, and it combines a wide range of variation with a high degree of interest. There are at least four types, with several possible

variations of each type. Type I consists of a two-column selection of a series of related facts; Type II consists of a means for testing reorganization of facts; Type III consists of a test for the regrouping of facts; and Type IV is a test where a multiple choice is offered for the selection of relevant from irrelevant facts.

Some types of the Selection Test do not lend themselves readily to the dictation method, although dictation of the entire test is possible in all cases. Administration by either the blackboard or the mimeograph method is more desirable, and it is probable that better results can be thereby secured.

In these tests there are special cases where general systems of scoring are impossible. When these types are used, care should be taken to select the proper method of scoring; and in these cases the methods recommended are especially advised.

The Selection Test is one that is easy to construct, one that is welcomed by pupils, and one that is easy and quick to score. It measures many aspects of school learning that are otherwise difficult to measure, and it can be used effectively as a method of review, as a method of teaching, or with equal success as a method of examining.

Sample Selection Tests ¹

SEVENTH-GRADE ENGLISH PAPER ²

RAW SCORE _____ DATE _____ M SCORE _____

This is a Selection paper. There is at least ONE error or omission in each of the following sentences. Correct or complete as the case demands. In Part II you will find rules which apply to your cor-

¹ These tests, constructed and used by Toledo teachers under the direction of the writer, suggest some of the possibilities in the use of the various types which have been described in this chapter. The first test is rather long, especially for use in a battery of tests, but it has certain diagnostic possibilities which are of interest.

² This test was constructed and used by Miss M. Beatrice Louy, McKinley School, Toledo.

rection or completion. In the margin of Part I write the letter name of the rule in Part II which best applies to your correction. Use each rule ONLY ONCE.

PART I

- 1. a stitch in time saves nine.
 ----- 2. Who was it said All that glitters is not gold?
 ----- 3. He asked me where I was going
 ----- 4. scott is the author of kenilworth.
 ----- 5. We shall go on tuesday.
 ----- 6. He cried, "O joy, i am saved at last."
 ----- 7. People came in great crowds from the north.
 ----- 8. The form of that play is somewhat different from the
 italian form.
 ----- 9. The items of the bill are as follows: apples oranges
 grapes bananas.
 ----- 10. new hampshire is east of vermont.
 ----- 11. John, i and James are going.
 ----- 12. Mary, John, you and I are to go.
 ----- 13. Little Boy Blue
 Come blow your horn,
 the sheep's in the meadow,
 the cow's in the corn.
 ----- 14. Considering that first, how should it be handled
 ----- 15. It was sent to Mr J L Smith.
 ----- 16. we remain,
 Yours very truly,
 Smith Gray and Co.
 ----- 17. Mr. Henry Irving
 my dear sir :
 ----- 18. childrens clothes are for sale here.
 ----- 19. That happened April 19 1775.
 ----- 20. Its a shame the bird broke its wing.

PART II

- A. Begin all proper names with a capital.
 B. Begin sections of the country with a capital.
 C. Words or phrases in the same construction forming a series should be separated from one another by commas.
 D. The period should be placed at the end of every declarative and imperative sentence.

E. Every direct quotation should be inclosed within quotation marks.

F. When the person spoken to is included with others, the person spoken to should be placed first.

G. Begin the first word of every line of poetry with a capital.

H. When including the speaker in enumerating others, the speaker places himself last.

I. Begin with a capital all adjectives derived from proper names.

J. The first word of the complimentary close of a letter should begin with a capital.

K. In contractions use the apostrophe to indicate omitted letters.

L. The question mark should be used after every sentence of direct question.

M. The period should be used after all abbreviations.

N. Begin the first word of every sentence with a capital.

O. Begin with a capital the names of the days of the week and of the months of the year.

P. The pronoun "I" and the interjection "O" are always capitalized.

Q. Begin names of states with a capital.

R. Begin the first part of the salutation of a letter with a capital.

S. The comma is used to separate parts of dates.

T. The apostrophe is used to denote possession.

Be sure that your name is on the back of each sheet.

SIXTH-GRADE GEOGRAPHY PAPER ¹

RAW SCORE _____ DATE _____ M SCORE _____

This is a Selection paper. Underline the ONE best ending in each of the sentences given below. The sentence must be true and make good sense. Be careful to do your best.

1. Much of the surface of the Philippines is covered with (grassland — jungle — desert — ice).

¹ This test was made and used by Miss Edna Roemer, Auburndale School, Toledo.

2. The Philippine Islands are under the control of (England — the United States — France — Spain).

3. The greatest product of the Hawaiian Islands is (rubber — apples — cane sugar — wheat).

4. The ocean voyage from New York to San Francisco is shortened by (Panama Canal — Straits of Florida — Suez Canal — Cape Horn).

5. One of the chief exports of the Philippines is (flour — hemp — meat — machinery).

6. The greatest industry of Alaska is (agriculture — mining — manufacturing — grazing).

7. Porto Rico is one of the islands of the (East Indies — Bermudas — Samoan group — West Indies).

8. The climate of the Hawaiian Islands is (hot — cold — changeable — mild).

9. The waters of Alaska abound in (perch — salmon — whitefish — bass).

10. Manila is the chief city of (Philippines — Porto Rico — Panama — Hawaii).

FOURTH-GRADE NATURE-STUDY PAPER ¹

RAW SCORE_____ DATE_____ M SCORE_____

This is a Selection paper. Underline the correct ending in each sentence given below. The sentence must be true and make good sense. Be careful to do your best.

1. "Tru-al-ly, tru-al-ly" is the song of the (meadow lark — song sparrow — bluebird — chickadee).

2. Ants make up more than half of everything eaten by the (flicker — killdeer — meadow lark — robin).

3. In color markings the downy woodpecker is much like his cousin the (red-headed woodpecker — flicker — cardinal — hairy woodpecker).

4. Because of his reddish-brown breast and blue back we can tell the (robin — blue jay — bluebird — barn swallow).

5. Of the sparrows the greatest pest is the (English sparrow — field sparrow — song sparrow — chipping sparrow).

¹ This test was constructed by Miss Hazel Scott and used by Miss Helen Boyles, Hathaway School, Toledo.

6. The one which says its own name is (barn swallow — song sparrow — killdeer — robin).

7. The bird that lays her eggs in other birds' nests is (wood thrush — bobolink — cowbird — catbird).

8. A member of the thrush family is the (Baltimore oriole — robin — cardinal — bluebird).

9. A nest that is cup-shaped with one or two stories is built by the (catbird — oriole — yellow warbler — bobolink).

10. This bird builds a nest in a hole in a tree (robin — hermit thrush — cowbird — downy woodpecker).

FOURTH-GRADE HISTORY PAPER ¹

RAW SCORE----- DATE----- M SCORE-----

This is a Selection paper. In the spaces in front of the numbers in Column I put the proper letters from Column II which best describe the people mentioned in Column I. Write clearly. Do as well as you can.

COLUMN I

COLUMN II

----	1. Pocahontas	A. Sailed around the world
----	2. Drake	B. Lived at Jamestown
----	3. John Smith	C. Visited Inca in Peru
----	4. Cortes	D. Discovered America
----	5. Pizarro	E. Sent colonists to Virginia
----	6. Columbus	F. Went to Mexico
----	7. De Soto	G. Found a great river
----	8. Raleigh	H. Was a little Indian girl
----	9. Balboa	I. Searched for the Fountain of Youth
----	10. Ponce de Leon	J. First saw the Pacific Ocean

Be sure that your name is on the back of this sheet.

¹ This test was constructed and used by Miss Beatrice Mathias, Pickett School, Toledo.

CHAPTER VI

THE ASSOCIATION TEST

Purposes of the Association Test. The Association Test gives to teachers an opportunity to test a large number of elements in certain types of school subject matter with greater rapidity and with greater accuracy than is usual with other types of classroom tests, and at the same time gives to the pupils an opportunity for self-expression such as is not provided by any of the other teacher's tests except the Judgment Test. It has many purposes and supplies many needs of the teacher. Because of its wide possibilities it is easy for the teacher to abuse the test and to make it extremely unacceptable to the pupils. Therefore it should be used sparingly and with full knowledge of the dangers involved. As will be seen, these dangers are largely concerned with a tendency toward concentration on unnecessary and even undesirable details, which in themselves tend toward unfairness and difficulty.

One of the prime purposes of the Association Test is to check rapidly certain phases of school instruction. There are many subjects in which it is difficult for the teacher to test in any adequate manner the many details of the work or to check the understanding which he is trying to create. The Association Test is of such character that it can be given to a group of pupils in a comparatively short space of time and check a wide range of subject matter. This becomes particularly valuable when the teacher is confronted with the necessity of covering a given amount of the curriculum in a given time. If the teacher can find those phases of the work which need added attention, if he can discover those phases where there is misunderstanding or misconception over the

wide range that has been studied, he has a starting-point for providing efficiently for the correction of the difficulties.

A second purpose of the Association Test may be thought of as a way of helping difficult portions of the work to be understood better. In subjects or phases of subjects where there are many new and strange words there is a strong tendency for pupils to acquire catch phrases or verbalisms. The more such phrases are used the more verbal they become and the further removed, in consequence, from realities. The Association Test provides an easy way for the teacher to check the meanings which pupils are attaching to these new or strange ideas and to provide for the early correction of the difficulties or for the additional help that is necessary for the better appreciation.

A third purpose of the Association Test lies in the field of technical additions to the vocabulary and ideas of pupils. Words or phrases and even ideas which are used in special or technical meanings and which may have common or non-technical meanings in addition are frequently a source of confusion and difficulty to pupils. The Association Test allows such difficulties to be discovered and corrected before the errors have had a chance to become relatively permanent, and serves as well to make stronger the correct impressions.

A fourth purpose consists in the emphasis which can be placed upon names, dates, and places, where such are important. It is felt that names, dates, or places are of relatively small importance in and of themselves, but that the associations which they may have are frequently of major importance in the understanding of school work. The Association Test allows such dates and names and places to be tested in a quick and efficient way, and at the same time it helps to locate the difficulties that have been encountered and the misconceptions that have been acquired.

A fifth purpose, which the Association Test shares with the Judgment Test, lies in the opportunities that it gives for the pupils to express themselves in a manner which can be

graded with fairness to all the pupils and which at the same time will reveal to the teacher the errors in his presentations or explanations of new materials. The test provides, therefore, for the protected testing of the more valuable elements heretofore tested in the traditional types of informal school examinations; yet at the same time it eliminates to a large extent the dangers and inequalities which under most conditions those traditional tests have exhibited.

Characteristics of the Association Test. As may be seen, the test has a wide range of usefulness, lending itself to history, geography, nature study, civics, and other school subjects of that character. Here the values of the work accomplished lie not so much in the field of increasing school skills, as is the case in arithmetic, spelling, and writing, but rather in the field of the broadening of knowledge or the changing of attitudes. It should be understood, moreover, that the Association Test does not test so much the range of knowledge or the changes in attitudes as it does the background of understanding and clearness as to facts which is so vitally necessary for the extension of knowledge or for changes in ideals.

The test presents to pupils in rapid succession a large number of separated and carefully selected key ideas or key words which, in the mind of the teacher at any rate, are intimately connected with the results which he has been trying to secure through his teaching. The words are flashed in rapid succession before the pupils, and only enough time is allowed for an association to be formed and recorded. The results should show rather conclusively to the teacher the success of his efforts, as well as the points where those efforts should be later supplemented with further work or explanation.

Construction of the Association Test. The construction of the test is perhaps as simple as that of any of the tests which have been described, although for best results the same care should be exercised.

STEP 1. SELECTION OF SUBJECT MATTER

The first step in the construction is similar to that of all other tests and consists in the careful selection of the range and character of the subject matter involved. In the illustration given here the subject matter consists of a wide range of the study of the geography of Canada in the sixth grade. The purpose of the test was to discover how well the pupils had made associations with new terms, names, and places, and with the specialized meanings of technical ideas.

STEP 2. SELECTION OF KEY WORDS OR PHRASES

The second step in the construction of this test is the selection of a number of the key words or phrases which should involve valuable associations with this study of Canada. These may be put down in any order in which they occur, and the number of selections may vary with the purposes of the test. As is shown in a later chapter, the number used in a battery of tests might well be no more than ten; though where the test is used alone, the number can be increased. The following were selected for this unit:

PRELIMINARY SELECTION OF KEY WORDS AND PHRASES FOR THE TESTING
OF A UNIT OF SIXTH-GRADE GEOGRAPHY

- | | |
|-----------------------|------------------------|
| 1. Maritime | 14. Newfoundland |
| 2. Selkirk | 15. Yarmouth |
| 3. Mackenzie | 16. "Bread-of-the-sea" |
| 4. Barren Lands | 17. Coal |
| 5. Chinooks | 18. Klondike |
| 6. Portages | 19. Dawson |
| 7. Hudson Bay Company | 20. Fox ranches |
| 8. Province | 21. Winnipeg |
| 9. Fredericton | 22. Soo Canal |
| 10. Ottawa valley | 23. Montreal |
| 11. Douglas | 24. Vancouver |
| 12. Pulp | 25. Quebec |
| 13. Conservation | |

STEP 3. PLACING UNITS IN CHANCE ORDER

It is possible that in the particular grouping of words as they are first written a word may, because of its sequence, provide an association for a word which follows or precedes it. The teacher may not himself be aware of the trend of association which influenced him to select the words, but it is entirely possible that such a trend may be in the form of connected meanings; so that it is safer, in order to be sure of definite association results, to place the statements in chance order in so far as it is possible to do so. Chance order may be gained as in the previous tests by writing on a number of slips of paper or cards the numbers of the elements, which are in this case twenty-five. These slips are then shuffled thoroughly and drawn in order, the first number drawn becoming the first element in the rearranged list, the second becoming the second element in the list, and so on for the remaining placements. In the case cited above one chance order as determined in this way might appear as in the following list, where the former numbers are given in parentheses and the new numbers in order.

REARRANGEMENT OF PRELIMINARY SELECTION OF TEST ELEMENTS
IN CHANCE ORDER

- | | |
|----------------------------|-------------------------|
| 1. (15) Yarmouth | 14. (7) Hudson Bay Com- |
| 2. (16) "Bread-of-the-sea" | pany |
| 3. (8) Province | 15. (3) Mackenzie |
| 4. (6) Portages | 16. (5) Chinooks |
| 5. (20) Fox ranches | 17. (4) Barren Lands |
| 6. (10) Ottawa valley | 18. (22) Soo Canal |
| 7. (19) Dawson | 19. (14) Newfoundland |
| 8. (1) Maritime | 20. (12) Pulp |
| 9. (25) Quebec | 21. (11) Douglas |
| 10. (9) Fredericton | 22. (21) Winnipeg |
| 11. (18) Klondike | 23. (13) Conservation |
| 12. (2) Selkirk | 24. (24) Vancouver |
| 13. (17) Coal | 25. (23) Montreal |

When all the numbers have been drawn and the test elements have been rearranged as shown above, the test is ready to be given in any of the ways outlined below.

Giving the test by dictation. The dictation method of giving the Association Test is one which the teacher will find very convenient. It is practically as easy to score and handle this test as to score and handle any of those that have been previously described. The pupils should all be given the same kind of paper and should have at least two well-sharpened pencils to insure no lost time because of breaking of points during the progress of the test. When the preliminaries are out of the way, the teacher might then give directions for the conduct of the test.

I want to find out how well you know a number of the new things which we have been studying about Canada. Take your papers and write your name near the top.

The teacher can also at this point give directions for getting any other information which is desired, such as date, grade, school, sex, age, and the like. The teacher should find out beforehand the number of available lines on each page, so as to know whether two or more sheets may be needed. If two sheets are to be needed, these should be distributed in the beginning, and the directions should include the writing of names on both of the sheets at the same time.

Now, when you have finished, turn your papers over and look at me, so that I shall know when you are ready to go on.

Here the teacher should pause to allow all the pupils ample time in which to carry out the directions and, when the form of the test is still unfamiliar, to allow time to help those pupils who may need it. He may then continue:

On the first line, near the left-hand margin, write the number "1."

At this point the teacher should diagram a page upon the blackboard and show just where the number should go, or should illustrate by holding up a paper, correctly arranged, like that which the pupils have on their desks.

Now skip a line and write the number "2" just under the number "1." Then skip another line and write the number "3."

When there are second sheets, it is important that all should begin with the same question number, as the teacher will later probably find it a help to separate the two sheets and score each one separately. Should this be the case, to have all the questions on both the first and the second sheet the same for all the class will make reference from first to second sheet or from second to first unnecessary and will thereby contribute to easier handling of the papers. The following direction is based upon the assumption that in this test two sheets are necessary and that each sheet will hold twelve or thirteen items comfortably. The teacher should vary the directions to suit his individual needs in this as well as in other matters.

When you have written the number "13," take your second sheet and write "14" in the left-hand margin on the top line. Then skip a line and write "15," and so on until you finish with the number "25."

Here again the teacher should pause until all the pupils have had a chance to complete the numbering of the two papers. The teacher can then start with the body of the test, dictating each unit as follows:

Are you all ready? I am going to read, slowly, twenty-five different words or phrases. In the spaces you have left on your papers you are to write a short, true statement about each of the things I read. Take your pencils and be ready to start. Don't copy the

words that I read. Just write your answers opposite the right numbers, and you may use both lines if you need them. Ready?

Number One: "Yarmouth." Think of some true statement about Yarmouth. Number One. Write.

Number Two: "Bread-of-the-sea." [Pause.] "Bread-of-the-sea." Number Two. Write.

Number Three: "Province." [Pause.] "Province." Number Three. Write.

This procedure may be continued for all the elements through 25. When this has been completed, the papers can be collected and passed to the teacher or exchanged among the pupils in the class for correction according to the method of correction which is to be used.

Criticisms of the dictation method. The dictation method is a good one to use, and for most purposes the teacher will find it quite satisfactory. It places upon the pupils, however, the responsibility of maintaining a correct numbering throughout the test, and it will be found that until the method of taking the test has been mastered by the pupils this responsibility will probably be a source of error. Another source of error, and one which the teacher should be careful to eliminate as far as may be, is the misunderstanding on the part of the pupils, owing to slow association or poor hearing. This, however, is of small account, particularly if the teacher speaks slowly and distinctly in addition to having previously made efforts to familiarize the pupils with his pronunciation and inflection. Pupils who are new to a school, and particularly to a teacher, however, frequently have difficulties of this kind; and where such pupils are included in a test group, when dictation is used, their papers should be given special consideration. If it is found that errors have probably been caused by faulty understanding, either the papers should be eliminated from consideration or the test should be given in another way.

Giving the test by the blackboard method. In the blackboard method of administration the process is similar to that of the dictation method, except that the words and their numbers are written on the blackboard instead of being dictated.

The preliminary directions with respect to paper and pencils, to the number of sheets to be used, to the writing of the identification on the backs of the sheets, and to the numbering of the sheets can be given in exactly the same way as in the dictation method. When the sheets are prepared, the teacher should, however, vary the directions to suit the new method.

I have written here on the blackboard under this covering twenty-five different words or phrases which have to do with our recent study of Canada. I want to find out how many of these you know ; so write a short, true statement about each of them on your papers. Be sure to write your sentences in the right spaces, and do not copy the words that are on the board. You may have —— minutes [the time allowed may vary at the discretion of the teacher]; and I will tell you when there are three minutes left, so that you can finish. Remember: Write a short, true statement about each of the words or phrases you see here on the board and do not copy the words. Just write the statements. Are you ready? Start.

Here the teacher should uncover the place on the blackboard containing the statements and allow the pupils to write. At the end of twenty minutes, or whatever other period the teacher has elected, he should say, "Three minutes more." At the end of that time the direction should be given, "Pencils up," and then, later, "Turn over papers." With that the administration is completed, and the papers may then be collected or redistributed according to the way in which the corrections are to be made.

Criticisms of the blackboard method. The blackboard method has the advantage over the dictation method that it saves the time of the pupils during the test period, which may be an important item in a crowded curriculum or in a school where the time allotment for each subject is small. It does not save the time of the teacher to any appreciable extent, however, although it is somewhat of a saving in energy.

As was intimated, the elements written on the blackboard should be covered until the time of the test, to prevent advance knowledge of the test elements on the part of the pupils. This has been described in a previous chapter. It introduces, however, a serious objection to the blackboard method, especially as it means the constant presence of the teacher in the classroom after the elements are written.

This method is more reliable, however, than the dictation method, from two points of view. It does not give as great possibility of error through misplacing of numbers as does the dictation method, although that of course is possible. The numbers are plainly written on the board with little chance of misreading, and they are closely connected with the words and phrases. In the second place, if the teacher writes clearly, there is less error by misunderstanding the words and phrases through unfamiliarity with the teacher. Misunderstanding may, however, be caused by inattention, in which case the problem of the teacher is different and the teacher must concentrate on problems of interest and readiness before being able to concentrate on the problems of learning and teaching.

Giving the test by the mimeograph method. Giving this test by the mimeograph method is, as in many of the previous cases cited, more satisfactory than giving it by any other method. It is a way of administering the test which involves more previous preparation than any of the other ways, but it is far more reliable and usually a greater help in teaching. It means making a stencil of the test elements and copying

from this stencil as many times as there are pupils in the class. The time taken in this initial preparation is, however, offset in some measure in the end by the greater ease of scoring and the better way in which the pupils can be shown their mistakes. There is little opportunity and little excuse for pupils to make mistakes in this form of test, as far as the numbering of the questions and answers is concerned, because the questions and answers are all together on the sheets during the entire operation of the test from the beginning until the class discussion following the test. This means that the reliability of the test is much increased over that of the other methods, since the errors which appear in the mimeographed form are errors, in all probability, arising from lack of knowledge of the test elements rather than errors resulting from faulty methods or misplaced numbering.

The first step in preparing the stencil form is to take the final series of statements as given on page 128, in this case giving them the numbers there found. A sample test form is shown below, which has been found satisfactory and which has yielded satisfactory results. It will be noted here that the general form is the same as shown in previous mimeographed examples and includes the test title, the date on which the test is given, spaces for the score totals, and page numbering. The directions here given should be especially noted, since they require a different form of answer from that in any of the tests previously cited. This form of direction is especially valuable because it tends to eliminate one of the greatest difficulties which has been found in this type of test. Some pupils are prone to give, wherever it is possible to do so, the definitions of the words involved, or else just a disconnected series of associated words. A definition is difficult to score, because one rarely knows whether it shows mere parrot repetition or true understanding. A set of disconnected words is difficult to score, though it may have real associative elements with the question given, because it is difficult to

trace the associations involved and more difficult to evaluate them when they have been traced. For a pupil to say that "maritime" means "near the sea" does not give a first-class answer to that particular question, though it is practically certain that if a pupil can give that type of answer he knows the technical meaning of the word in connection with the "Maritime Provinces" of Canada. So also with "bread-of-the-sea." If the answer given is "ocean — fish — dory — storm," it is again probable that the pupil knows pretty well the meaning of "bread-of-the-sea," but the teacher cannot be sure. The directions given below constitute the best protection against these two types of answers which has yet been tried by the writer and those who have worked with him.

In its final shape, ready for distribution to the pupils of a class, the mimeographed form of the test on Canada, cited above, might appear as follows :

Sample of Mimeographed Form of Association Test

SIXTH-GRADE GEOGRAPHY PAPER ¹

RAW SCORE _____ DATE _____ M SCORE _____

This is an Association paper. Write a short, true statement about each of the following. Think. Be sure you are right and do your best.

1. Yarmouth _____

2. "Bread-of-the-sea" _____

3. Province _____

¹ This test was constructed and used by Miss Laura Kuhr, Newton School, Toledo.

4. Portages -----

5. Fox ranches -----

6. Ottawa valley -----

7. Dawson -----

8. Maritime -----

9. Quebec -----

10. Fredericton -----

11. Klondike -----

12. Selkirk -----

13. Coal -----

14. Hudson Bay Company -----

15. Mackenzie -----

16. Chinooks -----

17. Barren Lands -----

18. Soo Canal -----

19. Newfoundland -----

20. Pulp -----

21. Douglas -----

22. Winnipeg -----

23. Conservation -----

24. Vancouver -----

25. Montreal -----

Be sure that your name is on the back of each sheet.

Scoring the Association Test. The Association Test may be scored in much the same way as was recommended for the Judgment Test. It involves four major operations: (1) the review of the answers to any particular question; (2) an

arbitrary gradation of the values for varying types of answers; (3) the application of these values to the actual answers given; and (4) the adding together of the total values gained on all the answers on any one paper in order to reach a final score.

In this type of test it will be found, as a rule, that the pupil either does or does not know the association that is desired. There will be differences in the values of the answers that are received, however, owing to the quality of the answer; but it is not likely that the gradation in value of the answers given will be as wide as in the Judgment Test, and for this reason it is unnecessary to provide as many possibilities in the score range. A range of from zero to 2 will usually be found sufficient and will, if the teacher uses careful discrimination, be satisfactory.

STEP 1. THE REVIEW OF THE ANSWERS

The first step in the scoring is to review rapidly the answers to a single question in order to determine the range of the actual answers given. Here the teacher will be materially aided if he will keep a record of the variety of answers and jot them down as they occur in the first sampling. A review of half the papers should reveal the majority of the probable answers given and would form a good basis for the determination of the gradation of values.

STEP 2. GRADING THE VALUES OF THE ANSWERS

The second step consists in the arbitrary determination of these values. Here the jotted answers which the teacher has transferred from the pupils' answer sheets should be rearranged in the order of merit, and the appropriate score values should be attached. All the answers having a merit of 2 should be classed together. All those of less merit or of doubtful merit might be classed as 1 in value, and those answers clearly wrong or totally ambiguous should be given

a score value of zero. It is unwise with these immature pupils to give scores of less than zero; so this type of score can well be left out.

STEP 3. SCORING THE PAPERS

When the score values have been assigned to the answers of varying merit, the teacher can begin the actual scoring of the answers on the pupils' sheets. This should be carefully done, and, until the teacher can remember the varying values without question, each answer should be referred to the teacher's sheet of score values for verification. This verification of values takes a little more time at the beginning, but it means that the scoring of answers is much more reliable in the end; and it also means the avoidance of the dissatisfaction which is likely to result when two pupils with the same answers compare notes after a test and find different values attached.

STEP 4. CALCULATION OF TOTAL SCORES

The final step in the scoring consists of the addition of the individual scores received on all the questions. When there are two sheets, each sheet for each pupil should be totaled separately, the sheets assembled after the totaling, and the scores on the second sheets transferred to the first sheets for the same pupils. This can be facilitated if the teacher is very careful to keep the sequence of papers in the two piles of first and second sheets exactly the same during the entire period of scoring. If this is done, which means merely performing each operation in exactly the same way with each pile, taking them up in order, and after each sheet has been scored placing them in the new pile of scored sheets in the same way, then the addition of the two scores for any single pupil is merely a matter of laying the two piles side by side, face up, and transferring the score on the second sheet to the first sheet for the same individual pupil and checking occa-

sionally for accuracy. The addition of the two sets of scores on the two sheets would constitute the total score of the pupil. When this has been completed, the papers are ready for handing back to the pupils. Before doing so, however, the teacher should make a record of the marks, or should handle them as is shown in the later chapters, to allow the necessary interpretations of the test results. It is usually best to do this before handing back the papers, though it is sometimes desirable to wait until the papers have been reviewed by the pupils, in order to check any mistakes which the teacher may have made. A pupil will be jealous of every point of credit which he receives and will be keen to find any place where he should have increased his score. As a result of this the teacher may use the class as a means of checking the accuracy and fairness of his marking, and thus wait until the papers have been discussed in class before making his interpretation of the test results.

This procedure, however, has one distinct disadvantage, in that the interpretations of the teacher are extremely valuable to the class as a whole and become more potent in their effects when presented at the same time that the papers are handed back, when the interest in the results is at a maximum. Unless a teacher has been very careless, moreover, in his marking, it is not likely that the corrections that could come out of the class discussion would change the interpretations of the teacher very materially, except in certain M-scale scores,¹ and these can be made after the class discussion as well as before it.

Disadvantages and values of the Association Test. The greatest difficulties of this test have been noted in a few instances above. One of them is that of uniformity of answers, such as definitions or catch phrases, which are generally well known by most of the pupils as a result of class drill or particular emphasis on the part of the teacher. These cannot be prevented in entirety, and it is not always wise to

¹ See M-scale scoring in Chapter XIV.

prevent them; but when a stereotyped answer is given by a large majority of the pupils, it is almost sure to be the result of memorization rather than true understanding. Memorization in and of itself is not at all bad. It is a good thing for pupils to memorize certain things; but when memorization takes the place of understanding on the part of the pupils and when the memorized phrases are mere catch words and nothing more, memorization is questionable. When school work becomes somewhat removed from the present needs of the pupils, as some of it does under the present conditions of our curriculum, this tendency to verbalism becomes readily apparent in this type of test, and the test itself may be used in a diagnostic fashion for locating those points where the teacher should make special efforts toward clearer and more living teaching.

A second difficulty arises, as has been said, when a pupil makes his answer a group of mere associated words. The difficulty here is that the teacher cannot tell with any accuracy the extent of the understanding of the pupil and cannot, therefore, score such answers with any surety. This type of answer can be prevented by making the directions for the test quite clear, by requiring a short, true statement rather than a series of disjointed words.

The values of the test are achieved when the purposes for which it was given have been realized. The greatest value, especially with the busy teacher, consists in the wide range of subject matter which can be tested in a comparatively short time, from which it is possible to get relatively reliable results to aid further teaching or review. For checking technical vocabulary and meanings, for testing new words and their usage, for emphasizing difficult portions of the work, and for fixing names, dates, and places, where such are of importance and where the associations therewith connected are of value, the tests are of great worth, unapproached by other types of tests in terms of the amount of effort and time expended in proportion to the range covered.

Chapter summary. The Association Test is valuable in testing wide ranges of subject matter in a relatively short time and is especially valuable in the nonskill subjects of the elementary-school curriculum. It presents to pupils a number of key words or key ideas, from which they are asked to write suggested short, true statements, explaining thereby their knowledge of the ideas involved. The steps in construction consist of the careful selection of the range and character of the subject matter which it is desired to test, the selection and construction of the list of key words or phrases above referred to, and the placing of these words in chance order for giving to the pupils.

The test may be given to the pupils by either the dictation, the blackboard, or the mimeograph method, although the last is preferable because it is more reliable and can be used with better results than the other methods described.

The scoring is relatively simple, but it involves care in achieving uniformity in the assignment of the score values, wherein lies much of the reliability of the score results. The teacher who uses this test in combination with other tests will find it a valuable addition to them; but if this test is used too exclusively by itself, it will undoubtedly have a tendency to encourage formalism and memorization.

Sample Association Tests

FOURTH-GRADE GEOGRAPHY PAPER ¹

RAW SCORE----- DATE----- M SCORE-----

This is an Association paper. Write a short, true statement about each of the following. Be sure you are right and do as well as you can.

1. Food -----

¹ This test was constructed and used by Mrs. M. W. Sheridan, Franklin School, Toledo.

2. Shelter -----

3. Cotton -----

4. Mississippi -----

5. Trading -----

6. Milk -----

7. Columbus, Ohio -----

8. Soil -----

9. Washington, D. C. -----

10. Australia -----

11. Erie -----

12. Plants -----

13. Mining -----

14. Ohio -----

15. Flour -----

16. Agriculture -----

17. Overland -----

18. East Indies -----

19. Clothing -----

20. Manufacturing -----

21. Sahara -----

22. Wool -----

23. Transportation -----

24. Wrangell -----

25. Atlantic -----

Be sure that your name is on the back of each sheet.

FOURTH-GRADE GEOGRAPHY PAPER

RAW SCORE _____ DATE _____ M SCORE _____

This is an Association paper. Write a short, true statement about each of the following. Be sure you are right and do as well as you can.

1. Pará _____

2. Antelope _____

3. Elephant tusks _____

4. Machetes _____

5. Tsetse fly _____

6. Kongo _____

7. Sahara _____

8. Nile _____

9. Nomad _____

10. Atlas Mountains _____

11. Mediterranean Sea -----

12. Suez Canal -----

13. Ship of the desert -----

14. Sand dune -----

15. Caravan -----

16. Statue of Liberty -----

17. Barley -----

18. Amazon -----

19. Brazil -----

20. Isthmus of Panama -----

Be sure that your name is on the back of each sheet.

CHAPTER VII

THE COMPLETION TEST

Purposes of the Completion Test. The Completion Test can be used to promote a number of purposes of the teacher and to do so in a way that is difficult in other forms of tests that have been described. Those who have used the test have found that it is usually more difficult for pupils than these other tests, and is sometimes not so eagerly welcomed. This may be because the test is difficult in itself, or it may be that the teachers who have used the tests have not been so careful as they might have been to make the tests fit the capacities of the pupils. When these tests are carefully made and when the pupils are encouraged to do well in them, they are a splendid means of measuring certain abilities of pupils.

The test provides for expression of a somewhat different sort from that found in the Judgment or Association Tests. Instead of having the pupil think out and write a whole sentence it is only necessary for him to provide a few words. However, the choice of words which the pupil makes indicates the degree to which he can express himself, and it indicates particularly the extent to which a grouping of familiar words suggests an idea that he should be expected to know.

A second purpose of the Completion Test has been accomplished in the way in which the test provides motives for the work of the pupils. After a trial or two a pupil finds the essential need of knowing well the ground which he has covered. He finds a real and immediate need for the results of his study, because he finds it impossible to take the test without that knowledge. This is not because of any game element or so-called "sugar-coating" or other external mo-

tive but rather from a compelling challenge to make the completions successfully just as soon as it is discovered that they can be made. The stimulation thus afforded is desirable and is in addition to any other stimulation or encouragement to do school work.

As a direct result of this stimulation to study, the Completion Test can help in promoting a better choice and use of words. It is soon found that words have definite and precise meanings, and that there may be more than one possible word that can be used in a given situation. With the development of this idea the pupils are encouraged to weigh varying values for different possible completion words and to make choices among them in terms of the different ideas which they represent. This in itself promotes literacy and provides a valid reason for pupils to think of the meanings of the words which they use in connection with the more specific ideas which the words represent.

In addition to these purposes the Completion Test provides an immediate reason for accurate and understanding knowledge. As has been indicated, the test provides a motive for the gaining of that knowledge, and it also provides a reason which can be appreciated and used by pupils in their study. Pupils soon see that when they have accurate and complete knowledge the test is easy to complete, but that when that accurate and complete knowledge is lacking the test is difficult if not impossible.

Characteristics of the Completion Test. The Completion Test consists of a number of true statements which are presented to the pupils with certain key ideas or key words missing. The plan of the test is for the pupils to provide, from the extent of their knowledge and from the degree of suggestion derived from the part of the statement that remains, the missing words. If the key words are chosen with discrimination and if there are real ideas involved, the test constitutes a real measurement of effort and ability, in almost every phase of elementary-school study.

Construction of the Completion Test. In order to make a Completion Test that will be neither too difficult nor too easy for his pupils, the teacher should follow carefully the steps given below.

STEP 1. SELECTION OF THE SUBJECT MATTER

The first step, as in all previous tests, is the careful selection of the subject matter which the test is to cover. As has been stated, the test in itself is difficult, and for this reason it is not wise to include too wide a range of subject matter, especially in the early tests. After pupils have acquired facility in taking the test and know what is desired, a larger range of subject matter is possible than at first. In the illustration here given the test comprised one of a battery of tests on the geography of the Pacific states as taught in a sixth grade.

STEP 2. SELECTION OF STATEMENTS

The second step is the selection of a number of sentences which are relatively simple and constitute truths contained in the range of subject matter selected. It is wise for a teacher to make these sentences quite simple and natural and not to anticipate the following step of the selection of the key words. That should be left until after all the statements have been selected, which will insure a group of clear and straightforward sentences with few traces of artificiality. The sentences as selected for this test are given below.

PRELIMINARY SELECTION OF COMPLETION SENTENCES

1. Agriculture is the leading industry of the Pacific states.
2. Where rainfall is scanty, irrigation or dry-farming methods can be used.
3. An enormous amount of standing timber is still found in Washington and Oregon.
4. Lava rock forms much of the surface of the Columbia plateau.
5. The Yakima valley is noted for apples.

6. The water traffic to the East has been increased since the building of the Panama Canal.

7. Salt, soda, and borax are products of the desert of California.

These statements indicate to a certain degree some of the larger truths which pupils might be expected to retain after a study of the Pacific states. In the first statement, for instance, the truth is revealed that in spite of the moving-picture industry centered in California (with which the pupils are in all probability familiar), in spite of the prominence which gold-mining has had in the history of the Pacific states and with which pupils are easily impressed, and in spite of the oil industry and the fishing centers that are so prominent there, agriculture remains predominant. So also in the second statement, which depends upon a knowledge of the changes in climate and rainfall of the section. The third statement states a truth with relation to one of the untapped resources of the region. The fourth indicates a knowledge of the character of one of the larger plateau sections. The fifth reveals some specific knowledge of one of the famous horticultural centers. The entire group of statements indicates a wider knowledge than is apparent merely from the facts involved on the surface, and one can be reasonably sure that this wider knowledge is possessed if the pupils can express the truths which the sentences themselves contain.

STEP 3. SELECTION OF KEY WORDS OR PHRASES

The third step consists in the elimination from these sentences of certain key words or phrases. Each sentence should be examined separately and only such eliminations made as will not detract from the intent but which will obscure the meaning. In the first statement there are three key ideas: *agriculture*, *industry*, and *Pacific states*. If the range of subject matter were wider than that embraced in *Pacific states*, these two words might well be eliminated.

Since in the case here cited, however, the range is confined to the Pacific states, no good end would be served by eliminating these words. If the word "industry" were eliminated, the sentence would be very easy to reconstruct, as the wording follows a common trend of thought. Where a very easy sentence is desired in order to develop confidence on the part of the pupils in their ability to make the completions, this word would be a good one to eliminate. However, in this case the key idea which measures best the residual knowledge of the pupils is contained in the word "agriculture," which for that reason should be the word eliminated from this sentence. The sentence would then read as follows (note the space indicating the missing word) :

1. — is the leading industry of the Pacific states.

The intent of the sentence is clear, to name the leading industry of the Pacific states; but the actual meaning of the eliminated word depends upon the pupil's knowledge, resourcefulness, and carefulness in the choice of words.

In the second sentence the idea of scanty rainfall, with the consequent choice of irrigation or of dry-farming methods, is the key idea of the sentence; and if these words are eliminated, a large range of pupil reasoning and pupil knowledge can be measured. This sentence might become any of the following, depending on the ease or difficulty which the teacher might be trying to impose :

2. Where rainfall is scanty, — or — methods can be used.

Here the intent is clear, and a definite clue is given in the first phrase. The sentence can be made more difficult by eliminating the word "rainfall" and making the meaning of the sentence even more obscure. The sentence might then read as below :

2. Where — is scanty, — or — methods can be used.

In the third sentence there are two key ideas, that of standing timber and that of definite location. Here the sentence might be constructed as follows:

3. An enormous amount of —— timber is still found in —— and ——.

In the fourth sentence the type of rock found on the Columbia plateau is the key idea, and the identification of that rock as well as the identification of the region might become the purpose of the elimination. The sentence as constructed might read as follows:

4. —— rock forms much of the surface of the —— plateau.

The fifth statement contains the key idea of the Yakima valley in connection with its fame for apple culture, and the sentence might read as below. To leave out the word "apples" might so increase the range of thought as to make the sentence valueless.

5. The —— valley is noted for apples.

The sixth statement has two key ideas, that of "water traffic" and that of the "Panama Canal." Either of them, however, depends upon the other; so both should not be eliminated. If, for example, "water" and "Panama Canal" are both eliminated, a logical result in the answers might be one concerned with land traffic and the part of railroads in its development. To prevent this and to focalize the answers it would be better to leave the word "water," though to do so makes the sentence quite easy to complete. It would then read as follows:

6. The water traffic to the East has been increased since the building of the —— ———.

To make this sentence somewhat more difficult and to obscure the intent still more it is possible to eliminate still

another word, "increased," which would make the sentence read as below. Under these conditions the sentence would require more constructive thought.

6. The water traffic to the East has been —— since the building of the —— ———.

The seventh statement has two key ideas, the three products mentioned and the desert of California, where they are found. Either but not both should be eliminated, depending upon the results which are desired. In this case it was thought best to insert the three products and leave the other idea for deduction, though the process could have been reversed with equal value. The final sentence then read as follows:

7. Salt, soda, and borax are products of the —— of ——.

STEP 4. CHANCE ORDER OF STATEMENTS

When the elimination of key words is completed, the next step, which the teacher may take if he feels it to be worth while, is the placing of the statements in chance order. This is not always necessary, though unless it is done the completed list of statements may reveal a connected thread of thought. This may or may not be good, depending on how much the thread of thought would contribute to the completion of the sentences by the pupils. In the following example of a part of a Completion Test the continuity of thought gave meaning to the statements, and a chance order would have been unwise.

1. In Pará everyone talks about —— (rubber).
2. To get the —— the trees must first be —— (sap, tapped).
3. It is then —— (hardened).

When the teacher feels that the test will be better if a chance order is made, he can get that order by placing the

sentence numbers on cards, shuffling the cards, and drawing them as has been described in previous chapters.

With the completion of these steps the test is ready for giving to the pupils in any way that the teacher may choose.

Giving the test by dictation. Dictation is next to impossible with the Completion Test and should be used only under exceptional circumstances. The reason for the difficulties will be discerned by any teacher who tries to dictate "Where blank is scanty, blank or blank methods can be used," or "An enormous amount of blank timber is still found in blank and blank." The statements lose much of their force in dictation, and in many instances the values are lost through the ridiculous ideas which are involved. When the dictation method must be used, however, the teacher should have each pupil copy the entire test, leaving the blank spaces for the missing words in his statements. Unless a pupil has an opportunity to cut and try, to fit different words in the spaces to see how they read and what they mean, and to see the entire sentence before him while he is making his decision, it is very difficult for him to make a wise selection.

Giving the test by the blackboard method. The blackboard method is a much more satisfactory way of giving the Completion Test. It enables a pupil to have the sentences before him all the time while he is attempting to fit the proper words into their places and to study the selection of the words and their effect upon the meaning of the sentences.

The teacher should write the sentences on the blackboard, leaving out the key words that were eliminated in the construction of the test. It is a wise plan to indicate each word by a space of constant length and to make separate spaces for each word when two words come together, as in statement 6, *Panama Canal*. The length of the blank space should not indicate the length of the word, because in some cases that might lead pupils unnecessarily astray. A line on the blackboard ten or twelve inches long for each key word

eliminated, regardless of its actual length, should be found satisfactory, and the explanation of the process should be made by the teacher to the pupils at the beginning of the test. He might introduce the test somewhat as follows:

Please take out paper and pencils and be ready to write a paper about the geography which we have just been studying. Write your names on these papers and then turn them over and be ready to begin after I have explained what to do. I have written on the blackboard a number of sentences in which there are missing words. This is a Completion paper. The size of the spaces that are left does not indicate the length of the words, but each space does indicate that there is one word left out. You are to fill in the words that you think make the best sense and are true. It is not necessary for you to copy down all the sentences that are on the board. Just write the word or words that are missing, and be sure to write the number of the sentence in the margin. Write down the number "1," like this [illustrating], and then read sentence 1 on the board. Decide what words fit in best and write them down in order opposite the number "1." When you have done that, write down, under the number "1," the number "2." Read the second statement and then write down after that number the words that you think are missing. You can do the same with all the rest of the sentences. I will give you twenty minutes to do them all [or any length of time that the teacher may choose].

It is a good plan to write the sentences on a portion of the blackboard where they can be covered with a map or a shade of some kind so that they will be hidden until the time of the test, and it is also necessary that the original writing be done at some time when the pupils are out of the room. This will prevent the character of the test from becoming known to some of the pupils.

Criticisms of the blackboard method. The blackboard method is much superior to the dictation method in that it allows the pupils to see the entire sentence at once and to judge the quality of their answers while they are seeing the sentence, but it has its drawbacks, which must be guarded against. One of these has been mentioned in previous chapters, the difficulty of writing the sentences on the board while all the pupils are absent and of keeping the sentences hidden while the pupils are in the room, until the time of the test. Another drawback is the chance of extraneous error on the part of the pupils, which creeps in unless the teacher is very careful to have the pupils number the statements and the answers on their papers correctly. If the teacher finds that there is any great tendency on the part of the pupils to misnumber their answers, it would be a wise plan for him to have the pupils write out the entire statements, underlining spaces for the missing words. The disadvantage in this lies in the amount of time which is virtually wasted by the pupils in writing a great deal more than is necessary. This is not so much a matter of concern in the upper grades, where skill in writing is rather firmly established, but it is a great consideration in the lower grades, where this skill has not been attained. Any time which the teacher can save in eliminating useless labor in these tests is time which may be put to a better use, and it is worth while for the teacher to have the pupils learn the value of careful numbering, even at the expense of errors in the beginning.

After the pupils have finished the test, the papers should be collected and passed to the teacher, when they will be ready for scoring. The names on the backs of the papers will furnish identification whenever it is desired, while the absence of names on the fronts of the papers will make it easier for the teacher to score them with a minimum of prejudice or bias.

Giving the test by the mimeograph method. The mimeograph method eliminates most of the difficulties of the

blackboard method and in addition makes it possible to give to the pupils a clearer reason for their mistakes, when the papers are returned to them. This is not so easy with the blackboard method, because it means that the statements must be copied again as they were at the time of the test. It requires, however, the making of a stencil sheet and the use of some sort of copying device; and since many schools are not yet equipped for this type of work, it is an impossible method in many places.

It will be noted, in the illustration which follows, that the title of the test and the date appear as in the tests previously cited. A space is also left for the final scores which may be given. The directions should be especially noted, since in this method of giving the test the complete directions are always of importance.

Sample Completion Test by the Mimeograph Method

SIXTH-GRADE GEOGRAPHY PAPER¹

RAW SCORE_____ DATE_____ M SCORE_____

This is a Completion paper. Fill in each of the spaces below with ONE word which will make the statement read sensibly and be true. The length of the space does not indicate the length of the missing word. Take your time and do your best.

1. — is the leading industry of the Pacific states.
2. Where — is scanty, — or — methods can be used.
3. An enormous amount of — timber is still found in — and —.
4. — rock forms much of the surface of the — Plateau.
5. The — valley is noted for apples.
6. The water traffic to the East has been — since the building of the — —.
7. Salt, soda, and borax are products of the — of —.

Be sure that your name is on the back of this sheet.

¹ This test was constructed and used by Miss Laura Kuhr, Newton School, Toledo.

Scoring the Completion Test. For the Completion Test the scores are made on a scale in which the values vary according to the quality of the answers. It is neither wise nor necessary, as a rule, to have the variation in score too great, but in most cases it will be found to be good practice to have a range of four possibilities, from zero to 3.

STEP 1. SAMPLING THE ANSWERS

The first step in the scoring is for the teacher to make a quick inspection of the actual answers, to determine the varying types of answers that have been given. Taking each statement separately on all the papers, he should jot down rapidly the different completions which are made, and these should then be rearranged in the order of their merit.

STEP 2. ASSIGNMENT OF CREDIT VALUES

All answers which carry the essential elements of the best answer, or are equivalent in both sense and resourcefulness to the best answer, should be given a credit of 3 points. Answers which have some merit but which are badly chosen should be given a credit of 2 points, whereas those which are ambiguous or doubtful in character should be given a credit of 1 point. The remaining answers, those which are entirely aside from the point or else make nonsense or direct untruths, should be given no credit. When this classification has been accomplished, the teacher is ready to proceed to the next step.

STEP 3. SCORING THE ANSWERS

The next step in the scoring is the application of the varying credits selected to the answers as they are given on the papers. It is a good plan for a teacher to use a red or a blue pencil in this operation, so as to distinguish his markings from those of the pupils. Each statement should be carefully read and the answers compared with the credit sheet

which was prepared in Step 1, and the appropriate credit placed in the margin for each answer. It is very important that the teacher practice absolute consistency in the marking of these answers, since the virtue of the marking is that every answer which has equivalent merit throughout the entire set of papers shall have the same credit.

STEP 4. ADDITION OF THE SCORES

The final step in the scoring of the papers is the addition of the scores which have been credited to each of the questions. This is a matter of simple addition, and the final score as received can be placed in any convenient place on the paper. It has been found convenient to place this score in the upper right-hand or left-hand corner and to turn in the corner of the paper after the score has been transferred to the teacher's record book. Then the score is the joint property of the teacher and of the pupil who wrote the paper. With the corner of the paper turned in, the score is not likely to be read by any of the pupils except the one who made that score.

Dangers in the construction of the Completion Test. There are several specific dangers in the construction of the Completion Test which the teacher should be careful to avoid if possible when making the test. As was stated in an earlier section, the Completion Test is frequently too difficult for the pupils for whom it is constructed. This may be caused by a too great elimination of words from the original sentences. In any context the following sample sentence would be of extreme difficulty, if merely because of the great elimination of words.

The —— River crosses —— from —— to ——.

This danger can be eliminated if the teacher will make a selection of the words to be left out, according to the directions given in the earlier sections of this chapter.

A second danger, however, lies in making the elimination too slight to provoke thought, as is shown in the following example :

The Yukon River crosses Alaska from east to —.

This danger can be avoided if the teacher will make an analysis of the character of the sentence and eliminate key words rather than just any words.

The third danger lies in the elimination of words of little consequence to the thought. The pupil is thereby led astray in an effort to find words of greater consequence in the sentence than is actually the case. The following illustration shows the result of this type of difficulty. The reader must remember that the general thought of this sentence is now familiar because of the previous examples, but that would not be the case if the sentence were a part of a test.

—— Yukon River crosses Alaska —— east —— west.

This danger can be avoided, too, by making a careful analysis of the statements and by making the elimination on the basis of words of real consequence to the thought in the sentence.

A fourth danger lies in the elimination of too many of the thought-provoking words. If the words of real consequence to the thought of the sentence are eliminated, it is clearly difficult for the pupils to get a set toward the proper completion of the sentence which is accurate enough to allow them to make a satisfactory completion. The following sentence shows the consequence of this type of elimination :

The Yukon —— —— Alaska from —— to ——.

There is nothing here which is likely to provoke any more particular thought of the Yukon River than of anything else which might have to do with the Yukon district, while the completion of the two final missing words will be entirely

predicated by the suggestion given in the words selected to follow the word "Yukon." When a completion problem has a large variety of answers of widely varying sense and meaning, the teacher can conclude that the particular sentence has had too many thought-provoking words eliminated. In such a case any fairly sensible answer is practically as good as another.

Chapter summary. The Completion Test offers to the teacher a convenient means of providing for expression on the part of his pupils without the expenditure of time necessary for a large amount of writing. It has been found to supply motives for the work of the pupils and also to contribute to the promotion of a better choice and use of words. The Completion Test also provides an immediate reason for accurate and understanding knowledge on the part of pupils, a reason which is frequently difficult for a teacher to furnish in the ordinary course of school routine.

The test consists of a number of statements which are presented to pupils by the teacher. These statements have certain key words missing, and the problem for the pupils is, from the extent of their knowledge and from their resourcefulness in accepting right suggestion and rejecting wrong suggestion, to fill in the missing words so that the statements are true and sensible. The test is constructed by making a number of statements and then by eliminating from those statements a number of the key words or key ideas which they contain.

The test is difficult to give by dictation, but either the mimeograph or the blackboard method will be found to be satisfactory.

The tests are scored on a variable basis with every like answer receiving a like score, and the total of the scores made by any pupil on all the questions in the test constitutes his final score.

There are certain dangers in the construction of the test, such as eliminating too many or too few words or eliminating

too many or too few of the thought-provoking words, but these dangers can be avoided by the teacher if suitable precautions are taken while the test is in the making.

Sample Completion Tests

THIRD-GRADE GEOGRAPHY PAPER ¹

RAW SCORE_____ DATE_____ M SCORE_____

This is a Completion paper. In each of the spaces in the sentences below put in a word to make the stories true. Think well and do as well as you can.

1. The Chinese belong to the —— race.
2. Eskimos have —— to draw their sleds.
3. The —— people wear wooden shoes.
4. Eskimo children have no ——, but they are told many stories.
5. In Holland the —— pump water from the lowlands.
6. The —— were the first to use tea.
7. The —— sit on the floor when they eat.

FOURTH-GRADE HEALTH HABITS ²

RAW SCORE_____ DATE_____ M SCORE_____

This is a Completion paper. Fill the blank spaces below with words that you think should be used there. The sentences must be true and make good sense. Do as well as you can.

1. We get the energy which we need for work and play from the —— which we ——.
2. We can help our bodies take care of the food which we eat by eating —— and at a —— time.
3. We should eat —— meals a day.
4. We should not eat cake, ——, ——, or —— between meals.

¹ This test was constructed and used by Miss Mary Gallagher, McKinley School, Toledo.

² This test was constructed and used by Miss Rose Clippinger, Jefferson School, Toledo.

5. We should eat some fresh, uncooked foods like — , — , and — every day.

6. To make our bodies strong we should eat food that contains such grains as wheat, — , or — .

7. Such foods as — , — , and eggs should be fresh when eaten.

8. The — which we eat should be made from — flour.

9. Fruits and — should be — before eating.

10. Working in a garden at home will give us both — and — .

FIFTH-GRADE GEOGRAPHY PAPER ¹

RAW SCORE----- DATE----- M SCORE-----

This is a Completion paper. Write in each of the spaces below ONE word which will make the statements read sensibly and be true. The length of the space does not indicate the length of the missing word. Take your time and do your best.

1. The Chinese eat with — instead of — .

2. Most of the trade goes through — , which is called the — of China.

3. — is the chief — of the people.

4. China changed from an empire to a — .

5. Nearly everybody in China flies — when there is a — .

6. Hongkong is owned by the — .

7. — is the capital of China.

SIXTH-GRADE NATURE-STUDY PAPER ²

RAW SCORE----- DATE----- M SCORE-----

This is a Completion paper. Fill in each of the spaces below with ONE word which will make the statements read sensibly and be true. The length of the space does not indicate the length of the missing word. Take your time and do your best.

¹ This test was made and used by Miss Edna Roemer, Auburndale School, Toledo. Attention is called to statement 5, which contains the ideas of "kites" and "holiday" but which was almost invariably filled in with the two words "away" and "earthquake," showing the influence of current events.

² This test was constructed and used in Toledo schools by Miss Valeria Humberstone, student in the University of Toledo.

1. The brown thrasher scratches among the —— for much of its ——.
2. The —— suspends its —— from the branches of a tree like a swing.
3. The meadow lark belongs to the —— family.
4. A song sparrow has a —— breast and —— its tail when flying.
5. The cowbird —— like a seal, has —— like a seal, and holds its —— up like a seal.
6. The —— is called by this name because it catches its food on the wing.
7. The tufted —— calls ——, ——, ——.

SEVENTH-GRADE ARITHMETIC PAPER ¹

RAW SCORE ----- DATE ----- M SCORE -----

This is a Completion paper. The problems below have been worked out and have some parts named. Fill in the names of the terms where they are missing. Write clearly and do as well as you can.

1. \$180 S. P.

$$\begin{array}{r} 150 \text{ Cost} \\ \$30 \text{ ——} \end{array}$$

2. \$150 Cost

$$\begin{array}{r} 30 \text{ Loss} \\ \$120 \text{ ——} \end{array}$$

3. \$150 Cost

$$\begin{array}{r} 0.06 \% \text{ of loss} \\ \$9.00 \text{ ——} \end{array}$$

4. Cost \$120 $\frac{0.08}{\$9.60} = 8\% \text{ ——}$ Loss

5. \$150 Cost

$$\begin{array}{r} 20 \text{ ——} \\ \$130 \text{ S. P.} \end{array}$$

6. \$150 Cost

$$\begin{array}{r} 1.06 \% \text{ of cost plus } \% \text{ of gain} \\ \$159.00 \text{ ——} \end{array}$$

¹ This test was constructed and used by Miss M. Beatrice Louy, McKinley School, Toledo.

7. \$230 Cost

180 S.P.

\$50 ———

8. \$200 ———

0.04 % of profit

\$8.00 Gain

9. ——— $\frac{0.04}{\$400) \$16.00} = 4\%$ of gain

Gain

10. % of profit $\frac{100}{0.08) 8.00}$ Profit

CHAPTER VIII

MAKING TRADITIONAL-TEST TYPES EFFECTIVE

Difficulties in making traditional tests effective. Enough has been said in earlier chapters to indicate that the traditional-test types of school examination are likely to be ineffectual as they are generally handled. In Chapter II the advantages and disadvantages of the traditional school examination have been discussed. There can be no doubt, however, that if proper measures are taken to prevent the disadvantageous elements in this type of test, it can be made an effective and useful instrument in measuring the success of pupils in certain elements of their school work.

In the previous discussion it was stated that the difficulties in the use of the traditional school examination are in the main two: first, the nonobjectivity in the scoring, and, secondly, the restricted range of the test itself. From what has been covered in the preceding chapters many other tests have been suggested to correct this latter difficulty, and through their use in combination with the more traditional forms the restricted range of the traditional form may not be a serious factor. The greatest difficulty that teachers have experienced in the use of the traditional type of school examination has been that of achieving objectivity in the scoring of the test itself. As has been pointed out, unless a teacher is very careful many elements may enter to influence that scoring, some of them proper and some of them irrelevant. Not only is this the case but it is also true that the relative weight which a teacher may place upon these influencing elements may vary considerably from one time to another and from one paper to another. One way, then, to help make this type of test effectual is for the teacher to

take such precautions as he may to turn this nonobjectivity in the scoring and this variability in the influence of scoring elements into elements which are, in fact, as objective as possible. To do this is difficult but not altogether impossible.

Selecting the objectives and the range of subject matter of the testing. As a preliminary the test should be so constructed as to decrease, so far as may be, the difficulty in scoring. This, of course, can be done most effectually with pure fact questions, but the true value of this type of test does not lie so much in this field, which is better measured in such tests as the True False, the Association, and the like, as in those fields less well tested or hardly tested at all by the tests which have been heretofore described. There is undoubtedly a large field of abilities, such as would be measured by test questions of the essay type beginning with "Discuss," "Describe," "Compare," and by certain types of "Why" questions, which would as a rule be measured only slightly by the tests which have been described. This is therefore a fertile field for the traditional examination and one that can be well measured if certain precautions are taken to make the questions as objective as possible, which, in turn, would tend to limit the range of subject matter included and increase the objectivity of the scoring.

In making out the test questions the teacher should first of all locate his objective in using this form of test, should decide which ability he wishes to measure, and should try to formulate his questions in that light. Is it to discover a certain process of reasoning, such as would be required to answer a question like the following? "Tell briefly why Chicago is located as it is." Or might it be to find out whether a pupil can describe a given condition, as would be illustrated by his reply to the following question? "Describe the wind movements as between areas of high pressures and areas of low pressures." Or might it be to reveal the degree of understanding of the pupil concerning a phase of history? "Compare Washington and Jefferson in their governmental policies."

As will be seen from the foregoing discussion, the decision as to the objective, once made, carries with it inevitably a certain delimitation of the subject matter. Therefore the teacher should be careful to make his questions specific, though not necessarily narrow, and avoid questions which would bring discursive answers on the one hand or "Yes" or "No" answers on the other. Questions such as "Trace the development of the Roman Empire from Cæsar to its fall" and "Who was the greater general, Washington or Cornwallis?" illustrate these two extremes.

Judging the length of the test and the relative value of its elements. After having decided upon the object of the testing the next step is to formulate the actual test questions. Here the teacher should bear in mind that, because of the greater ease of handling some of the other tests, the greater range which they can cover, and, as well, the greater ease of handling and interpreting the results, this test should usually be one part of a battery of tests. The object of the teacher should therefore be to equate it as nearly as possible in both length and difficulty with the other parts to be given at the same time. No rules, such as those given in the following chapter for use with other forms of tests, can help to accomplish this. Each teacher must decide this question for himself, measuring in terms of what he desires (and thinks he is likely to get) in the way of an answer in terms of the experience of his pupils (with which he should be familiar) and in terms of his own past experiences, particularly in these forms of tests. Though this is a very uncertain procedure and though it seems to be true that in general teachers are prone to underestimate the abilities of their pupils, as is shown by the greater number of pupils of over age than those of under age in our schools, nevertheless because teachers are accustomed to make and give these tests to their pupils they are likely to be able to gauge the pupils' difficulties better than are any others. They are probably thus enabled to judge fairly well these matters of length and difficulty.

Only so many questions should be included as are likely to balance, in possible score, in the relative value as a measure, and in the relative length of time that would be necessary to complete the questions creditably, any other unit in the battery of tests that is used.

Setting the standards for scoring. Up to this point, save in certain respects of the length and character of the test, the teacher will probably have done little more than he would have done under ordinary conditions. From this point on, however, the teacher should try to treat this test differently from what is usual.

In scoring the test the teacher should try to keep in mind the fact that is shown plainly in the treatment of the other tests that have been described, that the pupils' papers are to be scored not in terms of what the teacher has thought the pupils should have learned, nor in terms of what the teacher has thought that he has taught, but rather in terms of what the class as a whole has actually shown that it has learned, what the test reveals that the teacher has actually taught. This means that the teacher must try to remain impartial, that he must remain open-minded, and that he must be ready to change his opinions or his judgments in terms of what the class has actually accomplished.

The teacher, in order to make his scoring effective, should set up the standards by which he proposes to score the papers. Some of these standards can be objective, but others must be in terms of the judgments or opinions which the teacher holds. These judgments and opinions, however, should be in terms of the results which the teacher has tried to achieve, and these he can and he must decide for himself. Neatness, legibility, correct spelling, good sentence structure, and the like are always excellent classroom goals, and there is no reason why the degree to which they have been achieved should not enter as a part of the score in most tests, though teachers should remember that except in cases of English composition these matters should not be of paramount im-

portance. Such elements, however, as neatness in dress, success in past efforts, general attitude toward the class and the teacher, school record, age, sex, and race are matters which should be measured in terms of the test results and not directly. The teacher in his judgments as to the scoring standards of a test should keep such matters out, or let the test itself, on its face, determine their importance. If a "bad" attitude toward a teacher, if slovenliness in dress, if a lack of success in past efforts, or if sex or race or age has any effect whatsoever, it will be revealed on the face of the test and can there be treated objectively. It need not be treated and should not be treated on any basis of prejudice.

The objective standards for the scoring should have a prominent place, and these again should be decided by the teacher. The actual background of facts which the answers reveal is one phase which the teacher can consider; another is the type of organization of the answer; another might be the insight and resourcefulness shown by the pupil; and still another might be the extent to which the facts or organization or thought developed is pertinent to the topic under consideration. The teacher can add anything to this list which he himself conceives to be of value and which he wishes to measure.

Assignment of relative values to standards. When the teacher has decided upon *what* his standards are to be in any particular case, he then should assign for them relative values. Here, as in the scoring of the Judgment Test, the teacher should remember that what he is trying to achieve is justice for all in equal proportion. This assignment, therefore, should be made according to the teacher's best belief, in terms of his objectives; and if it is adhered to without discrimination, the injustice, if there is any, bears no more heavily on one pupil than on another.

Let it be supposed that the standards which a teacher has set up for a certain unit of a certain test are as follows:

1. Neatness, legibility, spelling, etc.
2. Organization of facts.
3. Validity of argument.
4. Pertinency and validity of facts.

The first question which the teacher should decide is the relative value of these four elements for the questions under consideration. In this case how valuable are neatness, legibility, spelling, and the like? Obviously if all that an answer showed was neatness and legibility, the answer would be worth but little. On the other hand, if neatness and legibility were low or nearly absent, the answer, no matter how good otherwise, might be worth but little more. Moreover, it should be clear that the first element should have less value than any of the other three, though what the relations of these other three might be would vary with the objectives of the teacher. In this case let it be supposed that the objectives of the teacher are in terms of validity of argument, supported by the pertinency and validity of the facts brought up in its defense. In this case, although organization would be important, it would be less so than the validity of the facts, and still less important than the validity of the argument. Thus from reasoning of this character the four standards would bear the following relation to one another: standard 3, *validity of argument*, should have the greatest weight and value; standard 4, *pertinency and validity of facts*, should have less weight; standard 2, *organization of facts*, should have still less weight; and the least weight of the four should be placed upon standard 1, *neatness, legibility, spelling, etc.*

The teacher should now assign actual relative values to these standards. If the question that is to be scored has a total value in the test of 5 points, these 5 points should be distributed among these 4 standards in terms of the general ratings decided upon above. If the number of points assigned to the question is greater than 5, if it is 10 or 12 or more, then that number of points must be distributed. This means

in effect that the teacher, before proceeding further, must know how many questions are to be included in this portion of the test and must divide among the questions in proportion to their relative value the total number of points to be given. In the case just cited let it be assumed that the teacher is using this question as a part of a Traditional-Type Test which forms a part of a battery of three tests, True-False, Completion, and Traditional-Type. Let it further be assumed that the teacher gauges the difficulty of the questions at about half that of the True False Test, which has twenty elements with a total possible score of 20. In order to equate the Traditional-Type Test with the other two units in the battery it should have approximately the same total score *for equal amounts of time spent upon them*, as is shown in the following chapter, which would give a total of two questions to the Traditional-Type Test, each question with a total value rating of 10.

On the basis of 10 points' value on the question under consideration, and in accordance with the relative ratings for the standards as they have been laid down by the teacher, the following might be a distribution of these 10 points in terms of the score standards:

1. Neatness, legibility, spelling, etc., 1 point.
2. Organization of facts, 2 points.
3. Validity of argument, 4 points.
4. Pertinency and validity of facts, 3 points.

With this as a basis the teacher should be in a position to score the test objectively, or at any rate more objectively than he could if he had not named his standards and delimited his values for them. It should be noted that the values assigned to these standards are positive values and that they are in terms of excellence rather than in terms of error. It should also be noted that these standards can vary in their elements as well as in the varying respective values of the elements, according to the purposes of the teacher

in assigning the test questions. Therefore each question, if necessary, can have a different set of standards and a different grouping of standard values.

Scoring the Traditional-Type Test. In scoring, the teacher should consider each question separately and judge each answer on each paper in the light of the standards he has previously set up. In order to get a judgment as to what the pupils have actually done in any question the papers should first be sampled and tentatively judged according to the standards set up. In the illustration cited the teacher could read a number of the papers and in doing so decide as to what he would consider creditable or noncreditable in standard 1, *neatness, legibility, and spelling*. In this case, where the limit of the score for this standard is 1, there is no reason, if the teacher can make fine enough discriminations, why part credits should not be given, such as $\frac{1}{2}$. It is not easy to be absolutely accurate in these discriminations, but having the standard would tend to decrease the total subjectiveness of the decision.

After determining standard 1, with its varying values, the teacher should turn to standard 2, *organization of facts*, and repeat the process. Here, from his sampling he should decide which of the forms of organization given should have a credit of 2 points, which of 1 point, and which of zero. With this in mind, and with, if possible, some notation to which he can refer or even certain samples withdrawn from the pile of answer papers to refresh his memory and define his types, he is ready to score the papers on this standard. With respect to standard 3, *validity of argument*, his samples should show several degrees of argumentation. These the teacher should try to classify in terms of his possible credit values, some as deserving 4; some, 3; others, 2; others, perhaps, 1; and still others, zero. He should then turn to standard 4, *pertinency and validity of facts*, and judge his samplings according to the values at his disposal. In each of his standards the teacher should try to reënforce and

maintain his opinion as far as he can through objective evidence. Especially in the beginning of the scoring of a group of papers these objective reminders, such as sample papers (which should be labeled plainly for ready reference), characteristic answers jotted down as in the Judgment Test, and, as in the case of neatness and the like, class standards which are well known, should be referred to, although it should not be long in the correction of a series of papers before these standards are remembered.

As each question is read by the teacher, the standards should be considered in turn and a judgment rendered with respect to each of them. As the weighting is valued, the score for that weighting should be written in the margin of the paper: (neatness) 1, (organization) 1, (argument) 3, (facts) 3. It is unnecessary to write in the particular item for which the credit is given, as these items can be explained in the review of the test which should come later. The only thing which would be necessary would be to place the three or four credits given in exactly the same sequence on each question. When the itemized scoring on any single question has been finished, the complete, or total, score for that question can be computed by adding up all the credits which have been received and writing down the total. A special indication, such as drawing a circle around the total score, will help to distinguish this from other score markings on a paper and will facilitate the final addition of scores for that portion of the test.

When one question has been completed, the scorer should pass on to the next and treat it in like manner, judging the varying quality of the answers, assigning credit values, marking these down, and computing the final score on the question. The last step in the process is merely that of adding together the scores on the individual questions to find the total score on the test.

This procedure may seem on casual reading to be somewhat pedantic or even useless. It is, nevertheless, as far as

its main aspects are concerned, exactly what the average teacher now does. All that it pretends to do is to focus certain elements in the teacher's consciousness and maintain the values that have been set up. Unless a teacher makes some definite attempt, on the one hand, to eliminate prejudiced and unsupported opinion in scoring and, on the other hand, to reënforce and maintain justifiable opinion, any fair or just marking of tests of this character is impossible; for the credits given will vary exactly in terms of the variation in the prejudices and unsupported opinions. The effectiveness of these tests depends largely on the consistency of the scorer, and that consistency cannot be maintained without definite and adhered-to standards supported by all the objective evidence available.

Uses and limitations of the traditional test. Some of the uses of the traditional test have been referred to in earlier sections of this chapter; namely, the possibility which it offers to measure certain abilities which are not measurable in like degree by any of the other tests which have been cited. There are, however, other uses of these tests, and likewise limitations beyond those of the other tests described, of which the teacher should be aware and of which the teacher should make use.

By using the total scores which are obtained in these tests the teacher can take advantage of many of the ways of using Teacher's Classroom Tests which are described in later chapters. In combination with the scores of other types of tests, or even alone, it is possible for the teacher to find a class distribution, to manufacture a class-distribution curve, to judge the quality of the test or the quality of its grading, and the like. It is also possible to use the total scores to find achievement ratings and ratios, as is later described, and to M-scale the test scores in the same way as in other tests. It is possible to use these scores for grading of various kinds, for promotion purposes, and for the classification of pupils in the ways that are described.

The test, however, has certain limitations that make it difficult sometimes to fit well with others of the Teacher's Tests. One of these limitations is that the total scores alone are of value in the methods that have been described; and those methods, save in limited form, that are based upon the scores on parts of questions or parts of tests, which can be located in True-False, Judgment Tests, and the like, are impossible with these tests. It is thus impossible to arrive at any useful analysis of question difficulty or to locate in any definite form the differences between teacher and pupil difficulties in the answering of questions.

Certain elements of diagnosis are of course possible through these tests; but the elements when found are somewhat doubtful, and the causes of the difficulties are more difficult to trace. It is always a question, for example, even when certain difficulties have been found, whether the difficulties are a matter of the pupil's lack of learning or a matter of the teacher's lack of teaching. The same limitation holds, as well, in any efforts to improve teaching, for the analysis which is given for this purpose in a later chapter is difficult to follow in the results which are received from these tests.

A serious limitation, however, in the use of these tests is in the weighting of the tests to make them comparable with the other tests that are used. Because so much of the weighting must be in accordance with the judgment of the teacher, and because there is no way of checking that judgment without an extended statistical procedure, and that after the test has been given, the real equality of these tests with others that are used is more questionable than it is with those others.

Chapter summary. The main difficulties encountered in making traditional-test types of school examination effective are two: the difficulty of making the range of the test comparable with the work which has been done by the pupils, and the difficulty in making the scoring objective. The first difficulty is minimized when these tests are used in

combination with others that are described in the previous chapters. The second difficulty can be reduced by first carefully selecting the objectives of the testing and the range of subject matter that is to be included; secondly, by calculating the length of the test and judging the relative value of the elements which make it up; thirdly, by setting up definite standards for scoring and assigning relative credit values to those standards for each question to which they are to be applied; and, fourthly, applying these credit values and standards in the actual scoring of the test papers.

The test, in its total scores, can be used for much the same purposes as can the other tests that have been described, but it is limited in its uses in such fields as educational diagnosis, improvement of teaching, and analyses of question difficulty, because the results of the tests in their specific elements cannot be used in the same way as the results from the previously described Teacher's Classroom Tests.

SELECTED BIBLIOGRAPHY

- MONROE, DE VOSS, and KELLY. Educational Tests and Measurements, Revised Edition (chap. xiv, "Improvement of Written Examinations"). Houghton Mifflin Company, 1924.
- LAIRD, D. A. "A Comparison of the Essay and the Objective Type of Examinations," in *Journal of Educational Psychology*, Vol. XIV, pp. 123-134.
- MONROE, W. S. Written Examinations and their Improvement. University of Illinois, *Bureau of Educational Research Bulletins* Nos. 9 and 17, October, 1922, and November, 1923.
- BRADY, J. An Investigation of the Written Examination. University of Pennsylvania, 1923.
- RUCH, G. M. The Improvement of the Written Examination, chap. iii. Scott, Foresman & Co., Chicago, 1924.

CHAPTER IX

THE USE OF BATTERIES OF TESTS

Advantages in the use of a battery of tests. A battery of tests consists of two or more of the preceding types of tests placed together and given at one sitting to a group of pupils. There are many occasions when a teacher would find it of advantage to use more than one of these informal tests at once. It may be that the making of one particular type of test will not be found to cover a range of subject matter in as thorough a way as the teacher would like, whereas if two or three types of tests were used the subject matter might be better tested and more variety given to the testing. It may be, too, that the teacher would like to test at one time more abilities than any single unit would test with accuracy, and thereby be able to have a more inclusive rating on the pupils of his class. In this case the use of two or three tests involving different abilities might solve the problem of the teacher and give a resulting measure of somewhat wider significance than if one type of test were used alone. The use of a group of tests frequently gives a teacher a greater opportunity for locating the difficulties which the pupils are experiencing in the teaching, and at the same time it gives to the teacher a greater chance to analyze his own achievements.

A further advantage which the teacher may find in using these tests in combinations is that they generally give a teacher a clearer notion of the relative standing of the members of a class. The importance of this phase of the use of batteries of tests is shown at greater length in a later chapter, where the use of the data resulting from this kind of testing is discussed.

From the point of view of economy of labor and time, too,

a battery of tests frequently is more easy to make than any single unit covering an equal range of material and requiring about the same time to complete. When a teacher is making a battery of tests and considering the materials from which it is to be made up, some portions may readily suggest statements for True-False elements, others may suggest Judgment elements, and others Completion, Selection, and the like. Thus two or three different units can be manufactured together and compiled as a battery of two or three different types of tests at little more cost of time or labor than any single test of less length.

In general it may be said that where a teacher is using these tests as a substitute for, or in combination with, the traditional type of informal school examination, in such school operations as the derivation of marks for promotion or school reports, a battery of two or three tests given at the same time will be found to be a somewhat better all-inclusive measure than any single test given alone. Moreover, it is not necessary for a teacher to eliminate entirely the traditional form of school examination, since it is possible to give it as one unit of a battery of these tests, as was described in the last chapter, and, in the subsequent operations which are described in later chapters, it can be treated in much the same manner.

Disadvantages in the use of a battery of tests. There are at least three major disadvantages in the use of batteries of tests, and the teacher should appreciate them. Where a battery of tests is used, the composite score of all three sections, or of the two sections, means little unless precautions have been taken to equate the scoring units of the several sections. While the scientific procedure for doing this is somewhat laborious and technical, this disadvantage can be largely eliminated in a less reliable but fairly satisfactory way, as is described in later sections of this chapter. A second disadvantage is that the use of a group of tests may seem to mean an increase of work for the teacher. This

is, however, not necessarily true; for, after a teacher has been accustomed to making these tests, it will be found that the various parts can be manufactured simultaneously with little more effort than would be required to manufacture a somewhat longer test of any of the single units used. For this reason it is also probably true that for equal range and equal amount of time spent a battery of tests is of greater value with less labor than any single test. In the third place, a battery of tests is likely to require more time, both on the part of the teacher and on the part of the pupils, but it is not likely to require as much time as the giving of two or three separate tests or, again, more time than any single test which would measure as much.

Process of making a battery of tests. The following steps in the making of a battery of tests have been found to lessen the difficulties that might be encountered and increase the efficiency both of the teachers and of the resulting tests.

STEP 1. SELECTION OF THE TEST TYPES

The first step in the process of making a battery of tests is the selection of the various types of tests that are to be used. It has been found that two or three types will usually give the best results, with three types to be preferred over two because more abilities are involved and therefore a somewhat wider range of markings is usually received. In the five general types of tests which have been described in the preceding chapters there are ten different possible combinations of three tests each which can be used. These are as given in Table I. It should be noted that if there is any reason for not using any one of the test types, there are four possibilities from which to choose. Thus, if a True-False form is undesirable, Group 7, 8, 9, or 10 is available; and if the Completion form is not wanted, the teacher can choose one from among Groups 4, 5, 6, and 9. The same holds true if any of the other three types is eliminated.

TABLE I. POSSIBLE COMBINATIONS OF THREE-TEST-TYPE GROUPS

1. True-False Judgment Completion	6. True-False Selection Association
2. True-False Association Completion	7. Judgment Association Completion
3. True-False Selection Completion	8. Judgment Selection Completion
4. True-False Judgment Association	9. Judgment Association Selection
5. True-False Judgment Selection	10. Selection Association Completion

If the traditional examination is included as a part of a battery of three tests, there are also ten possible combinations, where there are two of the newer test types included in each battery. These combinations are shown in Table II.

TABLE II. POSSIBLE COMBINATIONS OF THREE-TEST-TYPE GROUPS
EACH CONTAINING ONE UNIT OF THE TRADITIONAL-TEST TYPE

1. True-False Completion Traditional-Type	6. Completion Association Traditional-Type
2. True-False Judgment Traditional-Type	7. Completion Selection Traditional-Type
3. True-False Association Traditional-Type	8. Judgment Association Traditional-Type
4. True-False Selection Traditional-Type	9. Judgment Selection Traditional-Type
5. Completion Judgment Traditional-Type	10. Association Selection Traditional-Type

The selection of test types must be made with two or three considerations in mind. The teacher should consider first what he wishes to test and should adapt the test types both to that and to the materials at hand. A second consideration should be the way in which the test is to be given. If, for example, all three of the tests are to be dictated, no one of the three should be of the completion type. Again, if the test is one which is to be given in as short a space of time as possible, as well as being dictated in addition, it would be unwise to include either the Judgment or the Selection type, which usually require complete copying of the original matter before the test can be answered. This would mean that only a True-False Test and an Association Test should be given, which, in turn, would mean the lengthening of the test in accordance with Table III. It would probably be as wise for the teacher to give a part of the test in this case by the blackboard method, dictate the rest, and include three test types. In the selection of test types, however, the situation in which the test is to be used, the subject matter of which it is to be composed, and the means at hand for administering the test are all of moment.

STEP 2. EQUATING THE LENGTH AND VALUES OF THE TEST PARTS

The second step in the process of making the battery of tests is that of equating the length and values of the various parts of the test. As was stated in an earlier section, it is important that efforts be made to equate the various parts of the tests so as to make them in some measure comparable in their results.¹ It has been found that there are two simple ways in which this may be satisfactorily accomplished, both of which should be used. These are, first, to make, as nearly as possible, the length of time allotted to each part of the

¹ The illustrations used in Chapter XIV, in the section entitled "The desirability of composite scores," pp. 267-269, are applicable in this connection.

battery of tests the same, and, secondly, at the same time to make the total possible scores on each section as nearly the same as may be. Experience has shown that the following may be taken as basic units in this equating when the tests are fairly well adapted in difficulty for the group to which they are given and when the entire group of tests is given at one sitting.

Twenty True-False questions are about equivalent in the time necessary for their completion to that required by seven Completion elements, by ten Selection units, by ten Association elements, or to that required by seven Judgment questions. Thus if a test is made up of a True-False unit, a Completion unit, and a Selection unit, there should be twenty True-False statements, seven Completion questions, and ten Selection elements. This would equate the three parts of such a test, at least roughly, in point of the time necessary for them to be completed satisfactorily by the average pupils in a class.

In the scoring, if the general directions which have been laid down in the preceding chapters are followed, the tests, if given in the proportions shown above, will be automatically equated in their total scores. In the above illustration, for example, of twenty True-False statements, seven Completion elements, and ten Selection units the total possible score on the True-False section would be 20; on the Completion section, if the elements were scored on a basis of from zero to 3, the total possible score would be 21; and on the Selection section, if the elements were scored on a basis of from zero to 2, the total possible score would be 20. Thus for approximately equal time expended the returns in terms of point scores would have about the same possibilities.

There are more scientific methods of equating the units in a battery of these tests, but they involve statistical computations after the tests have been administered, with the resulting confusion in the score and the difficulty involved

in rescoring the papers after the relative values of the tests have been assigned.¹ In spite of the inaccuracies which it contains the way described above of making a rough equating of the parts of a battery of tests has been found worth while and accurate enough for the purposes of these tests. The following tables give in a more easily available form the information contained in the preceding paragraphs. Table III shows the number of elements, the score range for each of the elements, and the total score for each part of a test, according to the type of test that may be used in a two-type battery. Table IV shows the same information for use with three-type batteries. It will be noted that only the total score is shown for use with traditional types of tests, in accordance with the conclusions in the preceding chapter. It will also be noted that there are two possibilities of correction for Type III of the Selection Test.

TABLE III. NUMBER OF ELEMENTS, SCORE RANGE, AND TOTAL SCORE OF VARIOUS TYPES OF TESTS WHEN USED IN TWO-TYPE BATTERIES

TYPE OF TEST	NUMBER OF ELEMENTS	SCORE RANGE	TOTAL SCORE
True-False	30	0-1	30
Judgment	10	0-3	30
Selection — Type I	15	0-2	30
Selection — Type II	7	0-4	28
Selection — Type III	{ 15 30	{ 0-2 0-1	{ 30 30 }
Selection — Type IV	15	0-2	30
Association	15	0-2	30
Completion	10	0-3	30
Traditional	?	0-?	30

¹ One of these methods is that used in later chapters for reaching composite scores of different tests. This method of M-scaling the relative parts of a battery of tests is a good way of truly equating the parts; but it is cumbersome, and the method described above is satisfactory for all general purposes.

Another method for accomplishing the same end is described by McCall in *How to Measure in Education*, chap. ii, pp. 30-32, "Weighting tests according to the variability of their scores." The Macmillan Company, New York, 1922.

TABLE IV. NUMBER OF ELEMENTS, SCORE RANGE, AND TOTAL SCORE OF VARIOUS TYPES OF TESTS WHEN USED IN THREE-TYPE BATTERIES

TYPE OF TEST	NUMBER OF ELEMENTS	SCORE RANGE	TOTAL SCORE
True-False	20	0-1	20
Judgment	7	0-3	21
Selection — Type I	10	0-2	20
Selection — Type II	5	0-4	20
Selection — Type III	{ 10 20	{ 0-2 0-1	{ 20 20
Selection — Type IV	10	0-2	20
Association	10	0-2	20
Completion	7	0-3	21
Traditional	?	0-?	20

After the teacher has selected the two or three types of tests which he intends to use, he should refer to the table to determine the relative number of elements which each type should contain and should also determine from the table the range of credits which each set of elements should have in the scoring. Thus, if the group selected (8, Table I) consists of a Judgment, a Selection (Type I), and a Completion unit, there should be seven Judgment elements, each scored on a basis of from 0 to 3, ten Selection elements (Type I), each scored on a basis of from 0 to 2, and seven Completion elements, each scored on a basis of from 0 to 3. The total possible scores on each unit of the battery would be respectively 21, 20, and 21, or a total possible range of scores for the entire group of tests of from 0 to 62. When a Traditional-Type Test is used as one of the parts of a battery of tests, the equating of values and length between the parts should follow the plan outlined in Chapter VIII.

How to give a battery of tests. In giving a battery of tests to pupils it is not necessary to give all the test units in the same way. One part might be dictated, another part might be placed upon the blackboard, and another part might be mimeographed. As a general rule, however, the teacher will

find that if one part of the test is mimeographed, it would be as well to give all the units in that form; but for ordinary classroom use if the mimeograph method is out of the question, both the dictation and blackboard methods will be found satisfactory in any proportion which the teacher feels to be wise. No matter which method of administration is used, the special directions which are given in the preceding chapters are applicable, although the introductions by the teachers can be considerably curtailed.

Scoring a battery of tests. The scoring of a battery of tests should not involve any particular difficulties. Each unit of a battery of tests should be considered as a separate entity, and the special scoring considerations given in the preceding chapters may be followed for each type used. The only addition to the procedure consists in making a total score for the entire group of tests, which, if the equating of the units has been done in the ways suggested here, is easily accomplished by merely adding the total scores on the various units of the test for a final total score.

In the examples of tests which are given at the end of this chapter there is one which is scored and which shows how the final score was derived. It will be noted that in the mimeographed form the following is the score arrangement: Individual scores on the separate questions are written in the left-hand margin of the paper opposite the question numbers. These are added together for each unit separately, and the final total for that unit is written in the upper right-hand corner of the first page where the unit begins. When the unit begins in the middle of a page, the total unit score should be written in the right-hand margin at the beginning of the test section. This will serve to distinguish the separate scores. On the first page of the battery of tests the upper left-hand corner is reserved for the final total score of the entire group of tests. The scores should be transferred from the test units to their appropriate spaces in this corner and the

addition made directly there, giving the final total score for the test. Then, after the proper records have been taken, it is possible to turn in the corners of the sheets, which will conceal the scores until the pupils have had a chance to examine them.

One important consideration in the giving of a battery of tests is that the pupils should write their names on the back of each sheet of the test. In correcting the test units the teacher will find it of advantage to separate the various papers into the groups of like units, which will, of course, separate the sheets of the various pupils. It is a difficult thing to match up the papers after they have been separated, if the names of the pupils do not appear anywhere on them.

Another consideration that will result in easier scoring when the test is given in other than mimeographed form is to make sure that all pupils use the same kind of paper and that they place exactly the same number of answers on each sheet.

Chapter summary. A battery of tests consists of several units of the tests described in the preceding chapters combined into a single test group and all given at one time. In general the directions for giving the tests and for scoring the various units are the same as are given where the different types of tests are described. The additional changes which the use of a battery of tests brings about consist in the equating of the various test units by making the relative amount of time which is allowed for each unit approximately the same, and the total score allowed for the units approximately equal. Two short tables are given which show the proportions in which these tests should be combined and which also indicate the scores which should be allowed for the different types of test elements. The final score on the entire battery of tests consists of the total sum of credits which have been received on the separate units.

Sample Batteries of Tests

EIGHTH-GRADE HISTORY PAPER¹

Part I, 11

Part II, 14

Part III, 15

Total, 40

DATE-----

M SCORE-----

PART I

SCORE, Part I, 11

This is a Judgment paper. In the blank lines below write a short sentence which will tell the best reason you know why each of the statements is true. Do just as well as you can.

- (3) 1. On King Cotton the South based its hopes of success in the Civil War.

Answer. They expected help from other countries because other countries needed the cotton.

- (1) 2. Free public education is an American ideal.

Answer. Because it gives poor children a chance to go to school.

- (0) 3. Slavery prevented the South from becoming commercially prosperous.

Answer. Because the North did not have slaves.

- (3) 4. The Reconstruction period was more unbearable in the South than the war itself.

Answer. Because the government was in the hands of dishonest white people and negroes.

- (1) 5. "The reaper is to the North what slavery is to the South."

Answer. Because it did the work of the slaves.

- (0) 6. No state can lawfully get out of the Union. — LINCOLN.

Answer. Because there is no court of law for a state to go to.

- (3) 7. The people can be trusted to defend a government of which they form a part.

Answer. The people made the laws for themselves, and they believe them to be the best for all the people.

¹ These tests, together with the keys which are given on succeeding pages, were made and used by Miss Daisy Van Noorden, Lincoln School, Toledo.

PART II

SCORE, Part II, 14

This is a True-False paper. In the spaces in the margin below write the word "Yes" before each statement that you think is correct, and the word "No" before each statement that you think is not correct. Answer every statement. Do your best and THINK.

- (Yes) 1. Johnson proved to be a narrow-minded president.
- (No) 2. The slaves were a source of military weakness to the South.
- (Yes) 3. In the death of Lincoln the South lost her best friend.
- (No) 4. The negro ballot was desired by the South.
- (Yes) 5. Lincoln was in favor of dealing generously with the South.
- (No) 6. Secession was the cause of the War of 1812.
- (Yes) 7. Johnson put his Reconstruction policy into effect when Congress was in session.¹
- (Yes) 8. The Fourteenth Amendment gave citizenship to the negro.
- (No) 9. The "Underground Railway" system was formed by the Southerners.
- (No) 10. Grant succeeded Lincoln as president.
- (No) 11. Lincoln's plan of Reconstruction was adopted.
- (Yes) 12. The position taken by England pleased neither the North nor the South.
- (No) 13. Stonewall Jackson's brilliant campaign caused great rejoicing in the North.
- (No) 14. Fugitive slaves were constantly seeking refuge in the Confederate lines.
- (Yes) 15. John Ericsson built the *Merrimac*.¹
- (Yes) 16. The death of Lincoln occurred after the war was ended.
- (No) 17. Lincoln used emancipation as a military weapon.¹
- (Yes) 18. Grant's military ability won for him supreme command of the Union armies.
- (Yes) 19. Northern declaration in favor of freedom won them the favor of nations abroad.
- (Yes) 20. The close of the Civil War found the South almost ruined.

¹ Questions numbered 7, 15, and 17 on the True-False portion were wrongly answered.

PART III

SCORE, Part III, 15

This is a Completion paper. Write in each of the spaces below ONE word which will make the statements read sensibly and be true. The length of the space does not indicate the length of the missing word. Take your time and do your best.

- (2) 1. (*Gettysburg*), a (*great*) victory, was the turning point of the (*Civil*) War.
- (2) 2. Robert E. Lee, of (*Virginia*), was (*general*) of the (*Confederate*) Army.
- (3) 3. Nullification was advocated by (*Calhoun*) of (*South Carolina*).
- (2) 4. Johnson succeeded (*well*) and tried to carry out his (*Reconstruction*) policy but was not (*successful*).
- (3) 5. Slavery was first (*stopped*) in the District of Columbia, and the (*Federal government*) paid the owners for each slave freed.
- (3) 6. The decision in the (*Dred Scott*) case aroused the anger of the (*North*).
- (0) 7. "I made a vow," said Lincoln, "that if (*I*) drove Lee back across the (*line*) I would send the (*troops*) after him."

Be sure that your name is on the back of each sheet.

In the battery of tests reproduced above the answers to the various questions are as given by a pupil who took the tests. The corrections or values assigned by the teacher will be found in the margin, and the total scores for the various parts of the tests are in the upper right-hand corner of the page where each test begins. The total scores have been transferred to the first page of the test, and are there totaled to give a final test score, in this case a score of 40 out of a possible 62 points. The key by which this test was corrected and the various scores were assigned is given below. The values that are assigned to the answers are those actually used by Miss Van Noorden and represent her judgment of their relative worth.

KEY FOR EIGHTH-GRADE HISTORY BATTERY OF TESTS

PART I. *Judgment Paper**Element 1*

VALUES

IDEAS

- 3 Expected aid from foreign countries because of their need for cotton.
- 2 Had hopes of trading it with foreign countries for materials with which to carry on the war.
- 1 Cotton, the South's most important product.
- 0 Other answers.

Element 2

- 3 Equal opportunity the basis of democracy.
- 3 Educated people make better citizens.
- 2 Gives every boy and girl the same chance.
- 2 A country filled with ignorant people could not prosper.
- 1 Gives opportunity to the poor.
- 0 Other answers.

Element 3

- 3 The slaves were unfitted to work with machinery, but were well fitted to carry on agriculture.
- 2 Exchange of products with money invested in slaves left no money for large investments in factories.
- 2 The people were prosperous through cotton-raising and were satisfied.
- 2 Scattered population caused by the owning of slaves gave no opportunity for the building of commercial cities.
- 1 The people were absorbed in cotton-raising.
- 0 Other answers.

Element 4

- 3 Ruled by negroes and ignorant whites.
- 3 Lincoln's death and the harsh terms imposed by Johnson.
- 2 Government and property destroyed.
- 1 Poverty.
- 1 Difficulty in getting states readmitted into the Union.
- 0 Other answers.

Element 5

VALUES

IDEAS

- 3 Reaper did the work for the North as the slaves did for the South.
 3 It released the men from the fields, giving them an opportunity to go to the war.
 2 The South had slaves to do their farming. In the North the reaper took the place of slaves.
 1 Because it does what the slaves do.
 0 Other answers.

Element 6

- 3 No provision for withdrawal in the Constitution.
 3 Cannot withdraw without the consent of all.
 1 To be disloyal to the government is unlawful.

Element 7

- 3 The people made their own laws and think them right.
 2 Pride and work of building up our nation.
 1 They are defending themselves.
 1 Oath of citizenship.
 1 Loyalty.
 1 Interest.
 0 Other answers.

PART II. *True-False Paper*STATEMENT
NUMBER

ANSWER

1	Yes
2	No
3	Yes
4	No
5	Yes
6	No
7	No
8	Yes
9	No
10	No

STATEMENT
NUMBER

ANSWER

11	No
12	Yes
13	No
14	No
15	No
16	Yes
17	Yes
18	Yes
19	Yes
20	Yes

PART III. *Completion Paper*

1. Gettysburg — Northern, Federal, Union — Civil	3
Gettysburg — great, fine, wonderful — Civil	2
Gettysburg — Southern — Civil	1
Other battles	0
2. Virginia — commander in chief — Confederate, Southern	3
Virginia — general — Confederate, Southern	2
Other states — leader — Confederate	1
Other states — leader — North, Union, etc.	0
3. Calhoun — South Carolina	3
Anyone else	0
4. Lincoln — Reconstruction — successful	3
Lincoln — own — successful	2
well — Reconstruction — successful	2
well — own — successful	1
Anything else	0
5. Abolished, stopped — Federal government, United States	
national government, Union government	3
started and the like	0
6. Dred Scott — North, Northerners	3
Anything else	0
7. McClellan — Potomac, river — proclamation	3
McClellan — line — proclamation	2
Grant — Potomac — proclamation	1
Anything else	0

FOURTH-GRADE GEOGRAPHY PAPER ¹

PART I

[This is a True-False paper; dictated.]

- 1. The Amazon River is the largest river in South America.
- 2. Manaus is the most important rubber port of South America.

¹ This test was made and used by Mrs. S. D. Snow, Lincoln School, Toledo. The reader should note that although some of the elements in the Judgment portion of the test seem to answer some of the elements in the earlier parts of the test, because the test is dictated there is no chance for the pupil to change his earlier answers in case he should find himself wrong. If this test were to be mimeographed, the repetition would be undesirable.

- 3. There are many people living in the Amazon valley.
- 4. There are no fields and farms in the jungle land of South America.
- 5. There are no roads, except the rivers, in the jungle land.
- 6. The trees from which rubber is gathered are all of one kind.
- 7. The Amazon River flows into the Pacific Ocean.
- 8. The city of Manaus is west of Pará on the Amazon River.
- .. 9. The native Indians and negroes are the people who gather the rubber.
- 10. Large boats cannot sail up the Amazon River.
- 11. One would sail north in going from New York City to Pará.
- 12. There are many large cities in the Amazon valley.
- 13. Brazil owns more of the jungle land than any other country of South America.
- 14. The climate of any jungle is temperate.
- 15. There is little trade in the Amazon valley.
- 16. Rubber is gathered the year around.
- 17. The Portuguese are the rubber merchants.
- 18. The trees in the Jungle are of little value except for rubber.
- 19. Life in the Jungle is very different from life in Toledo.
- 20. South America is the only continent in the world on which there is a jungle.

PART II

[This is a Selection paper ; dictated.]

- 1. The ocean on which we would sail in going from New York to Pará is (Pacific — Atlantic — Indian — Arctic).
- 2. The largest river of South America is (Orinoco — Paraguay — Amazon — Paraná).
- 3. In going from Pará to Iquitos one would sail (north — south — east — west).
- 4. Tropical forests grow best in a climate that is (cold — temperate — humid — dry).
- 5. The city of Pará is noted for (sugar — wheat — cotton — rubber).

6. The country that uses the largest amount of the world's rubber supply is (France — Germany — England — United States).

7. The greatest rubber city in our own country is (Cincinnati — Chicago — Akron — Toledo).

8. The rubber is bought and sold by (Indians — Portuguese — negroes — French).

9. The greatest industry of the Amazon valley is (manufacturing rubber — farming — gathering rubber — making motor cars).

10. The most important product from the African jungle is (valuable wood — animal skins — rubber — ivory).

PART III

[This is a Judgment paper; dictated.]

1. There are few people living in jungle land

Answer -----

2. The rubber gatherers keep near the streams to tap the trees.

Answer -----

3. Rubber trees are now being cultivated on plantations instead of making use of all the trees of the forests that give rubber.

Answer -----

4. There are no farms in jungle land.

Answer -----

5. There are only certain seasons of the year in which trees are tapped for rubber.

Answer -----

6. There is little trade in the Jungle.

Answer -----

7. Horses, cows, and sheep are not raised in jungle land.

Answer -----

THIRD- GRADE NATURE STUDY PAPER¹

DATE _ _ _ _ _

PART I

[This is a True-False paper ; dictated.]

1. The flicker's tongue is very short.
2. The kingfisher sings a beautiful song.
3. The goldfinch is yellow with a black cap and black wings and tail.
4. The goldfinch has a white patch under his tail.
5. The bluebird is smaller than the robin.
6. The goldfinch is sometimes called the thistle bird.
7. The bluebird has a blue breast.
8. Mr. and Mrs. Flicker have red bands across the backs of their necks.
9. The kingfisher can run and hop.
10. The red-winged blackbird builds his nest in low bushes or marsh reeds.
11. The robin can hear the worms under the ground.
12. The kingfisher is larger than the robin.
13. The wings and tail of the flicker are yellow underneath.
14. Mrs. Red-Winged Blackbird has red and yellow patches on her shoulders.
15. The bluebird likes to build his nest in a hole in a tree or post.
16. The bluebird lives mostly in cities.
17. The goldfinch flies in a wavy line.
18. The goldfinch pumps the food down the throats of his babies.
19. The robin eats nothing but worms.
20. The robin likes to live near the homes of people.

¹ This test was made and used by Miss Marie Lerche, Sherman School, Toledo.

PART II

[This is a Selection paper ; blackboard method.]

COLUMN I

1. Kingfisher
2. Robin
3. Chickadee
4. Red-winged blackbird
5. Flicker
6. Nuthatch
7. Downy woodpecker
8. Goldfinch
9. Cardinal
10. Bluebird

COLUMN II

- A. A winter bird with a beautiful voice
- B. Comes down the tree headfirst
- C. Makes a feather bed for its young
- D. Likes ants
- E. Eats insect eggs from trees in winter
- F. Cousin of the robin
- G. Braces himself with his tail
- H. Plasters his nest with mud
- I. Has beautiful shoulder patches
- J. A fisherman

PART III

[This is a Completion paper ; blackboard method. Completions are in parentheses.]

1. The flicker belongs to the (woodpecker) family.
2. The (robin) sings many different songs.
3. The goldfinch (builds) his (nest) in July or August.
4. The kingfisher's nest is (in) the (ground).
5. The red-winged blackbird lives around the (marsh).
6. Sometimes the robin (runs) and sometimes he (hops) over the ground.
7. The (robin) and (bluebird) come early in March, at about the same time.

GENERAL BIBLIOGRAPHY FOR TEACHER'S CLASSROOM TESTS

- McCALL, W. A. "A New Kind of School Examination," in *Journal of Educational Research*, Vol. I, pp. 33-46, January, 1920.
- McCALL, W. A. *How to Measure in Education*, pp. 119-133. The Macmillan Company, New York, 1923.
- BRINKLEY, S. G. *Values of New-Type Examinations in the High School, with Special Reference to History* (Contributions to Education, No. 161). Teachers College Bureau of Publications, Columbia University, New York City, 1924.
- PATERSON, D. G. *Preparation and Use of New-Type Examinations*. World Book Company, Yonkers-on-Hudson, 1925.
- RUCH, G. M. *The Improvement of the Written Examination*. Scott, Foresman & Co., Chicago, 1924.
- BUTLER, W. F. "The Value of Informal Tests," pp. 94-119 of *First Year Book of Department of Elementary School Principals*. National Education Association, 1922.
- EATON, M. P. "New Style Examinations in Wadleigh High School," pp. 3-16 of *Bulletin of High Points*. New York, June, 1923.
- REMMERS, H. H., and Others. "An Experimental Study of the Relative Difficulty of True-False, Multiple-Choice, and Incomplete-Sentence Types of Examination Questions," in *Journal of Educational Psychology*, Vol. XIV, pp. 366-367.
- WOOD, BEN D. *Measurement in Higher Education*. World Book Company, Yonkers-on-Hudson, 1919.
- CUBBERLEY, E. P. *The Principal and his School*, pp. 486-488. Houghton Mifflin Company, Boston, 1923.
- TRABUE, M. R. *Measuring Results in Education*, pp. 475-479. American Book Company, New York, 1924.

PART TWO. WHY AND HOW TO USE TEACHER'S
CLASSROOM TESTS

FOREWORD

Iron ore, which is discovered and dug out by the prospector, must go through many processes before it can be used for rails, for bridges, or for buildings. It must first be heated and cast as pig iron, then remelted, changed to steel, and recast, and finally rolled, or hammered, or drawn, or recast, according to the specific uses to which it is to be put.

So it is with the results of these tests. The raw scores must go through certain processes designed to make them more fit for use and fit to prove their power. They must go through preliminary refining and casting, and then, according to their uses, — educational diagnosis (the skeleton of educational building), or improvement of teaching (the roadbed of education), or promotion (the bridges of learning), — these results must be hammered, or shaped, or drawn, or cast.

The earlier chapters which follow show the preliminary processes in the refinement of the raw scores, how and why to group the scores, how to interpret test curves, and how to locate question difficulties. Later chapters deal with the specific use of these results according to the specific needs of the teacher and the specific helps for pupils.

CHAPTER X

PRELIMINARY PROCESSES: THE DISTRIBUTION OF TEST SCORES

The need of a new point of view toward achievement. In the interpretation of the marks or scores which pupils receive on the Teacher's Classroom Tests the teacher must learn to regard marks from a different point of view and with different meaning from that to which he is accustomed in using the traditional school examination. In the traditional school examination teachers have been accustomed to the use of the percentile scale or its equivalent. For example, an examination of ten questions is given, and each question is marked on a 10-point scale. All the credits received on the examination are added together, and the total becomes, by unconscious conversion, the percentage mark which the individual receives. If the total is 70, then the grade given is 70 per cent, or a degree of correctness of 70 per cent as measured by the test.

These percentile grades, however, do not really form a measure either of an examination or of an individual, and it is unwise to use them with the teacher-made tests that have been described in the foregoing chapters. It would seem reasonable to suppose that pupils should be marked in one of two ways. They should be marked according to *their achievement in relation to their own ability* or else they should be rated according to *their achievement in relation to the ability of the group of which they are a part*. It is this latter type of marking which has been important in school work, because pupils are usually promoted according to the work which they have done in the group, and they are usually graded within the class according to what the class as a whole has done.

Pupils, should not, however, be graded on the extent of their achievement alone. It has been this which has been largely responsible, in all probability, for the tendency to cheat in examinations and to attempt to "pass," regardless of worth. The success with which pupils achieve results in terms of their abilities to achieve those results should be the criterion for rating. If, for example, a test were given to fourth-grade pupils and the same test were given to eighth-grade pupils, the eighth-grade pupils should be expected to do better than the fourth-grade pupils. In proportion to their ability, however, it should be expected that the fourth-grade pupils would do just as well as the eighth-grade pupils. So, if a score of 15 represents what might be expected of fourth-grade pupils and a score of 25 represents what might be expected of eighth-grade pupils in the same test, then, if a fourth-grader made a score of 15, it could be said that his achievement was consistent with the ability of fourth-grade children. However, if the eighth-grade pupil made a score of 15, it would be reasonable to suppose that his achievement was less than would be expected from a member of an eighth grade. Thus the fourth-grader who received a score of 15 might be entitled to an excellent rating, whereas the eighth-grader with the same score might be entitled to a failure. Whatever scale is used for interpreting the results of Teacher's Classroom Tests, it should be one which takes into account not only the total achievement of the individual but also the ability of the pupil or the ability of the group of which the pupil is a part.

The need of a fixed standard for interpreting achievement. From another point of view it is reasonable to suppose that pupils should be marked with relation to some fixed standard rather than with relation to a varying standard. The percentile scale used so frequently in the marking of the traditional school examination measures pupils' achievements by a varying rather than a fixed standard. The standard in this scale is determined by the difficulty of the test, and not

by either the ability or the achievement of the pupils. Some tests are easier than others. In fact, it is practically impossible for the teacher even to predict before an examination how difficult or how easy a given test may be. Let two tests be given, for instance, to the same group of pupils and marked on the percentage scale. Let the tests have also the same possibility for scores, — for example, a range of from zero to a possible 50, — and let the only difference between the tests lie in the fact that one is easy and the other difficult. On the easy test the pupils can do more than on the difficult test. The best score on the easy test might be 50, and the poorest score on the same test might be only 25. On the difficult test, however, the best score might be 25, whereas the poorest score might be nearly zero. The ability of the pupils has not changed. It is the difficulty of the test which has changed. A score of 2 on the difficult test, on the one hand, probably represents about as much real achievement on the part of the pupil as a score of 27 on the easy test. And yet in a percentile scale a score of 2 is a rating of only 4 per cent, whereas a score of 27 is a rating of 54 per cent, both meaning the same thing. If one is thinking, on the other hand, about the better pupils, in the easy test the best score is 100 per cent, whereas the best score in the difficult test is only 50 per cent. Yet the 100 per cent on the easy test and the 50 per cent on the difficult test mean exactly the same relative achievement.

If one is thinking of the tests as the center about which the pupils and the curriculum and the teachers revolve, then these marks are perhaps correct; but if one is thinking of the pupils as the primary source of educational effort, then one must think in terms of achievement in relation to ability. These ratings are then unfair. The standard by which they measure is in terms of the difficulty of the tests; and since the difficulty of the tests is variable, the standard is variable. It would seem clear that whatever scale is used for interpreting the scores on these teacher's tests, it should take

into account the difficulty of the tests and make use of a fixed rather than a varying standard. The pupils should not be penalized for lacking a knowledge which they cannot be expected to have, nor should they be penalized for the lack of ability of a teacher to equate the difficulty of his tests. They should be penalized in terms of what they might reasonably be expected to know and in terms of a test which is difficult in proportion to their abilities.

What pupils might reasonably be expected to know is determined by the pupils themselves. After all the teaching has been done, after the teacher has done all that he can to help his pupils to learn, the pupils should be measured in terms of what they have actually received and not in terms of what someone thinks they ought to have received. If the teacher has not done the teaching that might be expected of him, he is the one to be penalized rather than his pupils. For any group of pupils the average achievement of the group is a fixed standard, and all the pupils of the group should be measured by that standard. The use of Standard Tests will show whether that standard is too high or too low, and through the use of Standard Tests the proper remedies may be located and applied. The teacher may be stimulated to more or greater or different efforts, or some addition or change can be made in the curriculum, or some change may be effected in the methods or process of teaching which will bring the standard of the pupils to the standards shown by the universal tests to be desirable.

An analysis of the test used for determining just what pupils have received or gained will show wherein the teacher has failed or wherein the pupils have failed and will be useful as a guide for further teaching, and a grouping of the test scores will show the fixed standard that is desired.

The construction of a scale with a fixed standard. The following pages and chapters describe the details of a means for measuring by a fixed standard the results obtained in Teacher's Classroom Tests, at the same time equating the

difficulties of the tests. It corrects the two most serious objections to the use of a percentile scale. In brief the method involves, first, several preliminary processes for determining the achievement of the class and judging the worth of that determination; and, secondly, subsidiary means for using the results thus obtained, under the conditions outlined in the early sections of this chapter, in the tests for diagnostic and remedial purposes and in finding various pupil ratings. The first step consists in changing the class scores of a test into a frequency surface, which is dealt with in this chapter.¹ Other phases are considered in subsequent chapters.

The construction of a frequency surface. The pupils of a single grade, for the purposes of these tests, may be considered rather closely grouped about an average of all their respective abilities, and the standard set by the group as a whole may be considered the standard by which all the pupils should be judged. It is true that pupils are not yet, as a rule, properly classified within a school grade, but it is also true that they are constantly becoming better classified and that the method here considered is not seriously affected by the present situation. However, the better the classification of the pupils the greater is the value of these tests and, particularly, the more reliable are their results.

In order to find the achievement of a class it is necessary to find out how well the entire class has done. This makes necessary what is called a frequency surface of the class scores. The frequency surface may be made as follows :

¹ The usual statistical procedure is to make from a tally of the scores of a test a frequency distribution, or frequency table, and from that to construct the frequency surface. This is necessary when the number of individual scores to be handled is large, as would be the case in handling scores from all the pupils in a single school or from all the pupils in a single grade in a city. Where the number of pupils is small, however, — a single classroom of pupils at most, — the writer has found it economical of time and energy to make the frequency surface directly from the raw scores, as is here described. Reference, for those who desire it, is given at the conclusion of the chapter to standard texts where the construction of the distribution table is described in detail.

STEP 1. DETERMINATION OF THE TEST RANGE

The object in making the frequency surface is to put the individual scores of approximately the same size into groups which, when placed together, will give a picture of the accomplishment of the class. It is most convenient not to have too many of these score groups; so it is not wise to put into each score group only those scores of exactly the same size. If, for instance, the raw scores range from 15 to 61 points, with a possible maximum of 63 and a possible minimum of zero, as they are likely to do when two or three of the Teacher's Tests are used in a battery, it would take forty-six groups to contain all the scores if there were only one unit to a group.

TABLE V. SHOWING DIVISORS TO BE USED IN FINDING NUMBER OF UNITS FOR SCORE GROUPS IN FREQUENCY SURFACE ACCORDING TO NUMBER OF PUPILS IN CLASS

When number of pupils is 16 or less, divide range by 7
When number of pupils is from 17 to 20, divide range by 8
When number of pupils is from 21 to 25, divide range by 9
When number of pupils is from 26 to 30, divide range by 10
When number of pupils is from 31 to 36, divide range by 11
When number of pupils is from 37 to 42, divide range by 12
When number of pupils is from 43 to 50, divide range by 13
When number of pupils is above 50, consult reference below ¹

This is too many groups, because the resulting frequency surface would be a straight line which would tell very little of value. Therefore more than one size of score should be placed in each score group.

Because of the fact that interpretations of test distributions are valuable for the teacher, which is shown in the following chapter, the teacher should make the groups of such size as will give an easily interpreted frequency surface. Table V, given above, shows a quick method of determin-

¹ See W. A. McCall's *How to Measure in Education*, p. 361. The Macmillan Company, New York, 1922.

ing what should be the size of the score groups for a frequency surface. The teacher should therefore first determine *the exact test range in test points*. This is the range of scores from the lowest to the highest actually received by pupils on any test and is found by *subtracting the lowest score on the test from the highest score*. In the example given above, with the lowest score amounting to 15 points and the highest score amounting to 61 points, the test range is 46 points:

$$61 - 15 = 46.$$

With this range determined the next step is to consult Table V and to divide the range of score points by the number there found. If, for example, a teacher has twenty-seven pupils in his class, he should divide the score range, here 46 points, by 10, giving a result of 4.6:

$$46 \div 10 = 4.6.$$

This would be the number of unit scores to place in each score group. Since, however, the individual scores are in terms of discrete units, it is impossible to place 4.6 units in each group; so the nearest discrete unit should be used, in this case 5. According to the practice, then, in this example the frequency surface should be constructed with groups of scores covering a range of 5 test units each.

STEP 2. DETERMINATION OF THE BASE LINE

In the construction of this curve it is desirable for a teacher to use a cross-ruled paper with large squares or to rule a sheet of plain paper in squares about half an inch on a side. A line should be drawn near the bottom of this ruled sheet, which is called the *base line*. This line should cross enough squares to include the number of score groups that are necessary. In this case it should cross at least ten squares (10 is the divisor found in Table V), and the lower part of the sheet would appear as in Fig. 4 on page 208.

Beginning at the left-hand side the squares should be labeled with the scores of the various groups, with the lower scores placed at the left-hand side, as is shown in Fig. 5.

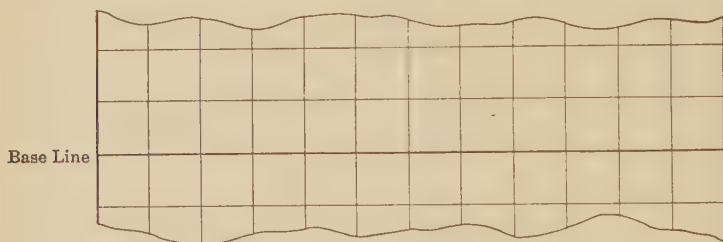


FIG. 4. Placement of the base line

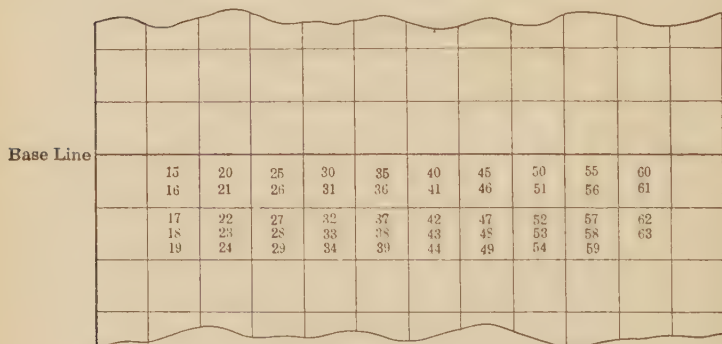


FIG. 5. Labeling score groups below base line

When this numbering of the squares below the base line has been completed, it is possible for the teacher to proceed with the construction of the frequency surface itself.

STEP 3. PLACING THE SCORES ON THE BASE LINE

The scores may be taken directly from the papers of the pupils, and each score put on the curve should occupy one square directly above the score group on the base line which contains that score size. The teacher will find it convenient

in later operations to place the exact score in the square, so that by placing the numbers in the proper squares the curve may be gradually built up without further markings.

Let it be supposed that the papers of a test have been corrected by a teacher and that the following are the raw-test scores as they appear on the several test papers:

40 — 32 — 29 — 48 — 38 — 37 — 51 — 23 — 34
 46 — 43 — 20 — 38 — 30 — 44 — 15 — 27 — 35
 25 — 61 — 49 — 37 — 42 — 32 — 34 — 40 — 39

The score on the first paper is 40. This number should be placed in the first square above the base line in the group numbered 40-41-42-43-44. It will appear as in Fig. 6.

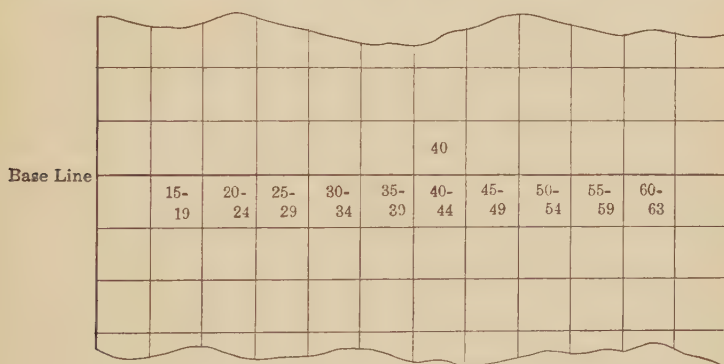


FIG. 6. Insertion of first score

In practice this score group, 40-41-42-43-44, is written 40-44, meaning "all scores from 40 to 44, both inclusive"; and therefore these score groups are written in this fashion in the succeeding illustrations.

The second score is 32, and this may be written in the first square above the group in which 32 is included, the group 30-34. The curve then appears as in Fig. 7.

The third score is 29, which is written in its appropriate square above the base line in the group 25-29. The process

is continued, as has been outlined, until all the scores as they are taken from the pupils' papers are written in their appropriate squares in the distribution. When a score appears where there has already appeared a score, the new number, even if it should be greater or smaller in size, should be written in the next square above the one previously filled in.

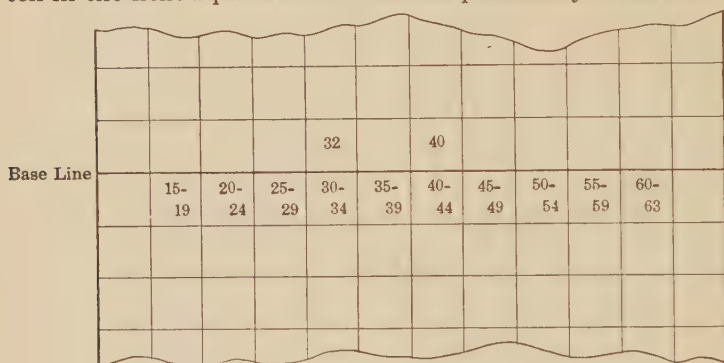


FIG. 7. Insertion of second score

In the illustration just cited, when all the scores have been written in their squares the surface will have an appearance as in Fig. 8.

When the curve is complete, a heavy line can be drawn around all the squares in outline, as is shown in Fig. 8, or else the squares that have numbers in them can be shaded. The result is a figure that is called a frequency surface. If the line is "smoothed," it is properly called a frequency curve. Either consists merely in a graphic representation of the test scores. It shows at a glance that in this test the scores group themselves rather evenly on each side of the group 35-39, that the individual who received the score of 61 was exceptionally superior to the other members of the class, and that the individual who received the score of 15 was evidently inferior to the rest of the class, whereas scores ranging from 30 to 44 might be considered as indicating the

class level. Further interpretations and uses of this frequency surface will be brought out in later chapters.

Chapter summary. The marking system of traditional school examinations is unfair because the percentile, or like, scale that is used takes account only of a total achievement rather than of an achievement with relation to ability and because the standard by which tests are measured on the scale varies, as a rule, with the difficulty of the particular tests that are used.

					39						
				34	37	40					
				32	35	42					
			25	30	38	44	49				
		20	27	34	37	43	46				
	15	23	29	32	38	40	48	51		61	
Base Line	15- 19	20- 24	25- 29	30- 34	35- 39	40- 44	45- 49	50- 54	55- 59	60- 63	

FIG. 8. The completed frequency surface

In the use of Teacher's Classroom Tests it seems wise, if possible, to correct these two difficulties caused by using the percentile scale. Accordingly a method has been devised for transmuting the scores made on Teacher's Tests into ratings that measure the achievement of an individual in relation to the achievement of the group of which he is a part, or in relation to his own ability, and which also make correction for the difficulty of the tests used, thereby employing a fixed rather than a variable standard.

The first element of this method consists in the manu-

facture of a frequency surface of the scores made in any test. In making this it is necessary to find first of all the range of the test scores, and to follow that with a determination of the number of score units which should be placed in each score group. The further steps consist in placing the scores in proper position on a cross-sectioned sheet and in outlining the final surface that is thus obtained.

The interpretations of varying kinds of curves and the use of the frequency surface are described in later chapters.

SELECTED BIBLIOGRAPHY

- MCCALL, W. A. *How to Measure in Education*, chaps. xii, xiii. The Macmillan Company, New York, 1923.
- TRABUE, M. R. *Measuring Results in Education*, chap. x. American Book Company, New York, 1924.
- GREGORY, C. A. *Fundamentals of Educational Measurement*, pp. 273-275. D. Appleton and Company, New York, 1922.
- RUGG, H. O. *Statistical Methods applied to Education* (chap. iv, Part IV, "The Frequency Distribution: The Steps in its Construction"; and Part V, "The Graphic Representation of Educational Data"). Houghton Mifflin Company, Boston, 1917.
- PATERSON, D. G. *Preparation and Use of New-Type Examinations*, pp. 70-71. World Book Company, Yonkers-on-Hudson, 1925.

CHAPTER XI

PRELIMINARY PROCESSES: THE MEANING OF CURVES

Curves have meanings which may be interpreted. With the completion of the frequency surface or the graph of the scores of the tests the teacher can reach some conclusions with respect to the general meaning of the curve which has been constructed. If a line be drawn around the limits of the squares so that a "smoothed" curve results, a judgment as to the quality of the test as a whole may be reached. This judgment will help not only in understanding why pupils reacted as they did to the test but also in making future tests. In Fig. 9 on the following page the curve as formulated in the preceding chapter has been smoothed for purposes of interpretation.

After smoothing the curve the teacher should extend the base line of the curve on the left-hand side down to the point which would represent a zero score, and on the right-hand side should extend the base line up to that point which would have been reached had some pupil received full credit on each element in the test. Note the extension on both sides of the curve in Fig. 9. This extension is important, as will be shown, for in interpreting a curve not only the general shape of the curve but also the relative position of the curve on the base line must be taken into consideration. The following illustrations will serve to show some interpretations of varying general types of curves, when it is assumed that the pupils of a class fall, with respect to general ability, into a more or less normal distribution. In a normal distribution there are a few who are poorer than all the rest, and an equal number who are markedly superior. The rest of the group,

in increasing aggregation, gather rather evenly about the average of the entire group. Such a grouping is approximately illustrated in the curve given in Fig. 9 and in Fig. 10.

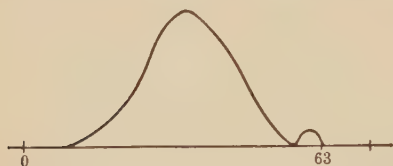


FIG. 9. Smoothed curve of Fig. 8

When a teacher has a group that differs markedly from this type of distribution, the following interpretations are not likely to hold in the same proportion. The directions given in Table V of the preceding

chapter, however, will serve to develop curves of the character here shown, the only difference among the curves with different sized groups of pupils being the size of the curve itself. The general shapes remain much the same for the same characteristics of the measurement.

In the following pages several general types of curve are illustrated, and general interpretations of the varying types are given; following each interpretation is a short paragraph dealing with the changes which might have been made in the test from which the curves were derived, which would have resulted in improvement and which might lead to better future tests.

TYPE 1. THE NORMAL DISTRIBUTION

When the distribution of scores on the test takes the form shown in Fig. 10, or one similar to it, the teacher can be assured that the test as devised and given has the following characteristics.

It will be noted that the curve rises from a point above the zero point of



FIG. 10. The normal distribution

the base line with a rather gradual curve, which increases as the curve rises higher, gradually decreasing in abruptness

as it reaches its highest point, and then falls to the base line near the high point on the scale.

Some interpretations of a test giving such a curve might be as follows:

1. It was a fair test, because it contained elements which tested even the poorest pupils as well as elements which extended the best pupils to their limits. In addition it contained a set of elements so graded in difficulty as to measure all the other pupils adequately.

2. The test was marked correctly, because a sufficient number of points, or credits, was allowed on each question to discriminate adequately among the variety of answers which was received.

TYPE 2. MINUS-SKEW DISTRIBUTIONS

Form A. When the curve takes the form shown in Fig. 11, which is called a minus skewed curve, the following interpretations may be made.

Note that this curve begins somewhere in the lower half of the base line, gradually rises to some point above the mid-point of the



FIG. 11. Minus skew—Form A

base line, and then drops abruptly to the base line as shown.

The teacher can draw the following inferences from this type of curve:

1. The test measures the poorer half of the class fairly well; so it can be assumed that the easy parts of the test were adequate.

2. On the other hand, the test did not measure the upper half of the class nearly so well, as is shown by the abruptly dropping line of the curve, which indicates rapidly increasing difficulty in the test. It can therefore be judged that the difficult part of the test was too difficult for the best pupils. An analysis of the question difficulty, as shown in the next

chapter, will reveal to the teacher the particular parts of the test which were too difficult, and from them the teacher can discover whether the fault lies with the questions, with the teaching, or with both. If the fault lies with the questions of the test, this test would have been improved if the more difficult questions, which the curve shows were too difficult, had either been changed so as to be more in keeping with the true ability of the pupils or else had not been marked with such great severity.

Form B. A second form of minus-skew distribution may appear much as in Fig. 12. It will be noted here that the curve has much the same general form as the preceding



FIG. 12. Minus skew—Form B

curve, but that it is farther to the right on the base line. The lower part of the curve starts somewhere near the center of the base line and gradually increases

until it reaches its greatest height, when it breaks abruptly downward, touching at or near the upper limits of the base line.

This curve has a different interpretation from that of the preceding curve and should be compared with it.

1. That there were no low scores on the test is an indication that the easy portions of the test were too easy and were completed without difficulty by all the pupils. It is evident that there were many questions in this test which the teacher might have taken for granted and which were therefore needlessly asked.

2. Not only were the easy portions of the test too easy, but the difficult portions of the test were almost entirely absent. There were in the test no questions or, in any event, too few questions which were difficult enough to measure the better pupils in the class. As a result nearly half the class found the difficult portions of the test so easy that they received nearly perfect scores.

As a result of a distribution of this sort the teacher has

a choice of judgments: either that the pupils have been subjected to most exceptionally excellent teaching or else that the test was not difficult enough to measure the pupils, or perhaps a mixture of both. In all events the teacher should in future tests endeavor to make them more difficult and have the test form a better measure of the capabilities of the pupils. The chief difficulty here would seem to be that the teacher had underrated the abilities of the pupils.

Form C. A third form of the minus-skew distribution is shown in Fig. 13. Here it will be noted that the curve starts near the lower end of the scale, rises gradually to a point near the upper end of the scale, covering nearly the whole of the possible range of the scale, and finally drops off abruptly near the point of its greatest height, as is shown above.



FIG. 13. Minus skew — Form C

The interpretation of this curve may be as follows:

1. The lower half of the class is fairly well measured, since the ranking of the pupils is well distributed. It is evident, however, that these easier test elements have too wide a range. This may be due to the fact that there are a large number of questions in the test that have approximately equal difficulty, or it may be due to the fact that too much emphasis has been placed upon certain portions of the test and that they have been given too many credits. The presence of the first element may be determined as shown in the following chapter, and the second element is explained in Chapter IX on "The Use of Batteries of Tests."

2. The upper half of the class has very little distribution, as is shown in the sharp drop of the curve at its upper end, and many of the pupils have achieved nearly perfect scores. This is an evidence that the upper reaches of the test were not of sufficient difficulty for the better students and that, from this point of view, the test should have been extended or else made to include some more difficult parts.

A test which shows this sort of distribution would have been improved by grading the easier elements of the test more carefully and by either being more severe in the marking of the sections where the teacher's judgment entered into the marking or else including a lesser number of the easier questions. The test would also have been improved if some more difficult questions had been included or if a more difficult type of test had been used as one of the parts.

Form D (undistributed zero scores). When the scores pile up at the zero end of the scale, as shown in Fig. 14, and very



FIG. 14. Minus skew—Form D

few pupils are able to make any score at all, and none of them are able to do much, it is evident that the entire test has been too difficult and is beyond the range of ability which the pupils possess. No such test should be used in the com-

putation of grade scores for individual promotion or for any other purpose than to correct the conceptions of the teacher as to the abilities of the pupils.

TYPE 3. PLUS-SKEW DISTRIBUTIONS

Form A. One form of plus-skew distribution which a teacher may find as a result of the giving of a Teacher's Test is shown in Fig. 15.

The curve rises abruptly from or near the zero end of the scale until it reaches its height, when it drops gradually to a point near or slightly above the mid-point of the base line.

An interpretation may be made as follows:

1. This curve indicates that the easier portions of the test were much too difficult for the poorer pupils, because so many of them were unable to make any adequate score. Even the pupils of medium ability have been able to accomplish little.

2. The more difficult portions of the test were also too difficult for the better pupils, because none of them was able to reach a very good score on the test.

When a curve of this kind is obtained, it is an indication that the teacher has overrated the abilities and accomplishments of his pupils. The correction would have been to

eliminate the larger portion of the most difficult questions and to substitute an equal number that were of an easier type, adding them to the remaining questions.

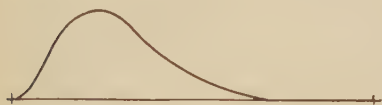


FIG. 15. Plus skew — Form A

The test would then have been in a form more suitable for the group to which it was given.

Form B. A second form of plus-skew distribution is illustrated in the curve shown in Fig. 16.

Here it will be noted that the curve starts considerably above the zero point of the scale, or perhaps even near the mid-point of the base line, rises quite abruptly to its maximum, and then drops gradually to a point near the high end of the scale.

An interpretation of this test might be as follows:

1. The fact that all the pupils were able to do a large portion of the test satisfactorily is an indication that the easier portions of the test were too easy even for the poorer pupils.

This means that there were too many easy questions included and not enough questions of medium difficulty.

2. The curve, however, indicates that the more difficult

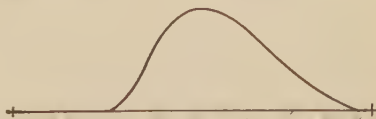


FIG. 16. Plus skew — Form B

portions of the test were satisfactory, because the better pupils were adequately measured. It also indicates that the more difficult portions of the test were rather skillfully adapted to meet the abilities of the better pupils.

This would have been a better test if fewer very easy

questions had been included or if many of these questions had been increased somewhat in difficulty or if the easier questions had been more severely graded.

Form C. A third form of plus-skew distribution may be found in many cases substantially as shown in Fig. 17.



FIG. 17. Plus skew—Form C

In this case the curve rises abruptly from a point at or near the zero point of the scale and then gradually recedes to a point at or near

the upper end of the scale. At no point except at the lower end does the curve rise very far above the base line.

An explanation of this curve might be as follows :

1. This curve indicates that the easier portions of the test were too difficult for the poorer pupils of the class, because so few of them made any progress on the test.

2. The results also indicate that although the test contained elements which measured the better pupils in the class, there were too many of these elements; and if they had been better graded, just as good results would have been secured.

The teacher may make a judgment with respect to the value of this test to the effect that it would have been improved if there had been a larger number of easier elements included in the test or if, perhaps, there had been included as one part a form of test which would have been somewhat easier for the pupils to do. At the same time the test would



FIG. 18. Plus skew—Form D

have been improved if the difficult portions of the test had been somewhat decreased in number without materially altering their range of difficulty. This result might perhaps have been accomplished if the teacher had scored one portion of the test with somewhat less severity.

Form D (undistributed perfect scores). When the scores pile up at the upper end of the curve, as is shown in Fig. 18, and very few of the pupils have done less than the maximum that was possible, it is an indication that the entire test was much too easy. Such a test, too, should not be used for any purpose other than to correct the teacher's judgment with respect to the abilities of the class. It is useless for purposes of promotion, diagnosis, classification, or grading.

TYPE 4. MULTIMODAL CURVES

The multimodal curve may exhibit many forms or types from rather distinct sets of modes to merely indistinct irregularities. The merely irregular curve should cause a teacher no great concern in its interpretation. It simply means that there are gaps in the gradation in difficulty of the questions, a matter which even the best examiners have difficulty in avoiding. It is only when the multimodal curve shows in great distinctness, as in Fig. 19, that it has great meaning, and then only in proportion as the various modes are prominent. The teacher should know that multimodality may be caused by taking too small measurements in the construction

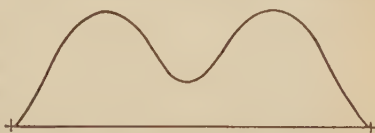


FIG. 19. Multimodal curve

of the curve, which is the case when True-False Tests that occur in score groups of two-point intervals are distributed in one-point units. It will be found that every other score group will be blank, thus forming a very irregular curve. In this case making the score groups contain at least two units will serve to make the curve more compact and easily read, even if it does make the number of score groups less than the number required by the table in the preceding chapter.

The most frequent case of extreme multimodality which a teacher is likely to encounter is illustrated in Fig. 19. Here

there are two distinct modes: one at the lower end of the scale and one at the upper end of the scale, with a gap between them. The curve rises abruptly from the lower end of the scale near the zero point on the base line and falls rapidly almost to the base line near its mid-point, then rises abruptly again, falling almost as abruptly to a point near the high end of the scale.

The curve might be interpreted as follows:

1. This curve shows that the easier portions of the test were too difficult for the poorer students, because so many of them cluster about the area of low scores.

2. The curve also shows that there were few questions of medium difficulty, because the pupils who were able to answer the more difficult easy questions were able to answer other questions of no greater difficulty until they finally reached the limit of their abilities in the more difficult questions, which they were unable to do at all. Many of these pupils received higher scores than they really should have been entitled to, had the test been better graded in difficulty.

3. The curve shows also that the very difficult questions were too easy for the best students, because so many of them were able to achieve high scores.

This test would have been improved if the easier questions on the tests had been more carefully graded in difficulty, with a greater number of moderately difficult questions added and with several much easier questions included. It would have been further improved if some of the more difficult questions had been made easier and if some of the most difficult had been made harder. The chief trouble with this test is the fact that only two types of question were included in the original make-up, and that there was a great gap in difficulty between them. These two types of question were of moderate ease and of moderate difficulty.

General observations on interpretations. The interpretations given above for the various forms of curves are of course dependent upon the fact that the abilities of the class

are distributed in a nearly normal curve. The teacher can determine this distribution fairly well from an analysis of the results of some good Standard Test which has carefully graded difficulties or, preferably, from the composite of a number of such tests. Attention is here called to the section devoted to the use of Standard Group Intelligence Tests for determining standard M scores, as given in Chapter XV. If the "raw scores" (not Intelligence Quotients) on such a test are put into a frequency surface, a fairly exact basic curve for judging a test curve can be obtained. If, for instance, a teacher finds that his class is, in fact, a skewed group and that the normal abilities of the class assume in a distribution some such shape as appears in the skewed curves given above, he must modify his interpretations to suit his particular case. If the group had a frequency surface in ability like that of Fig. 12, for example, the most desirable test curve would have the same general shape, and therefore the interpretations for such a curve, as given for Fig. 12, would not hold.

In general it may be said that when a curve rises sharply from the base line or falls sharply to it, it is an indication that there is a rapidly increasing difficulty in the questions of the test where the rise or the fall takes place. On the other hand, when the curve rises gradually or falls gradually, it is usually an indication that the gradation in the relative difficulty of the questions is well chosen. If the base line of the curve, from the zero point to the point of perfect score, be divided into four equal parts, those parts of the curve in the lowest quarter may be considered as resulting from the easy-easy questions. If the curve rises sharply in that section of the base line, it is then an indication that the easier questions are too hard. If the curve in that section rises gradually, it is an indication that the questions are well graded. If the curve does not start at all in this section, but rises from the base line at a point nearer the mid-point of the line, it is an indication that these easy-easy questions are too

easy. The four groups of questions might be named as follows:

1. The easy-easy questions.
2. The difficult-easy questions.
3. The easy-difficult questions.
4. The difficult-difficult questions.

By comparing the frequency surface with the graph of question difficulty, which is described in the next chapter, it will be easy for the teacher to identify those questions which have been largely responsible for the shape of the curve that has resulted from any test. An analysis of this sort will be of great help in the making of future tests.

Further modifications in the interpretations resulting from any test must be made, also, if the tests happen to be given at a time when the whole class is not present. If several of the best pupils or if several of the poorest pupils are absent when such a test is given, before the interpretations can be held as valid the teacher must make allowance for the change that such absence would make in the normal curve of the group.

Chapter summary. This chapter has shown that the shape of the distribution curve is of great importance in the interpretation of the character of the tests that are given. It also has shown that from the curve obtained a teacher is enabled to make some judgments which serve both as a check on the scoring judgments and as a very distinct aid to the teacher in understanding the reactions of the pupils. Many of these judgments are not of great importance in respect to the test then under consideration, but affect more largely tests which the teacher may make in future times for the same pupils. The systematic interpretation of curves of this character which are obtained through the actual testing of pupils will aid a teacher very materially in constructing future tests of more value and of better quality for both his present group and future groups of pupils.

General interpretations are somewhat dangerous, although it may be said that when the curve starts far up on a base line it is probable that much of the test is too easy for even the poorer pupils, whereas if it ends far down on the base line the most difficult parts of the test are probably too difficult for even the best pupils. A sharp rise in a curve probably indicates a rapidly increasing difficulty in a test, and a sharp fall has the same indication. Either a gradual rise or a gradual fall probably indicates a gradual increase, and a desirable increase, in the question difficulty. A teacher should make every effort to interpret such curves as he may obtain and should follow this interpretation with an analysis of the actual difficulty of the questions, as is explained in the following chapter.

CHAPTER XII

PRELIMINARY PROCESSES: THE DETERMINATION OF QUESTION DIFFICULTY

The importance of a question analysis. The preceding chapter has shown that one of the important factors in understanding the results of a test is an analysis of the question difficulties. This will give the first clue to the reasons which actuated the pupils to make the mistakes that they did make. It will furnish a firm grounding for the diagnostic work, especially with the poorer pupils, and it will furnish the teacher with definite data upon which to revise either the materials, the presentation, or the methods of teaching which have been employed. The first step in these processes of diagnosis and improvement of teaching consists in the analysis of the question difficulties.

The analysis of question difficulties is a most interesting process. It is easy, and because of the definiteness of the test scores, the results lend themselves to tabulation or to graphic presentation, which is frequently more emphatic than mere tabulation. Where just one type of test has been used, such as a True-False, a Judgment, or a Completion Test, it is possible to rank all the questions from easiest to most difficult and to show the relative differences in degree of difficulty of the various elements. Where, however, more than one type of test has been used in a single battery of tests, the ranking of all the elements becomes a more difficult procedure because of the differences in the numerical rating of the different sets of questions. For this reason the determination of question difficulty will be discussed under three different headings: first, the determination of the question difficulty of True-False Test types; secondly, the determination of the

question difficulty of variably scored test types; and thirdly, the determination of the question difficulty for batteries of tests where more than one type of test is used.

1. THE DETERMINATION OF QUESTION DIFFICULTY IN TRUE-FALSE TESTS

The True-False Test presents a relatively easy means for the determination of question difficulties. When all the papers have been scored and the scores have been carefully checked, the teacher should prepare a question-difficulty tally sheet for the test.

STEP 1. PREPARATION OF A QUESTION-DIFFICULTY TALLY SHEET

This sheet is merely a tally sheet for determining the number of times that each question was answered incorrectly by all the pupils. The tally sheet should contain a series of numbered lines, this series corresponding to the number of questions contained in the True-False Test.

STEP 2. TALLYING THE NUMBER OF INCORRECT ANSWERS TO QUESTIONS

The test papers should be piled in front of the teacher so that they can be easily read, and beginning with the top sheet the numbers on the sheet which have been crossed or checked as incorrect should be read and tallied on the tally sheet. If the first paper shows questions numbered 1, 7, 8, 15, and 17 as incorrect, tally marks should be placed opposite these numbers on the tally sheet. When one paper has been completed, the tally should be continued with the next paper, the new tally marks being added to those previously set down. The final result should show exactly how many times each question was missed by all the pupils.

TABLE VI. TALLY SHEET FOR A TRUE-FALSE TEST (TWENTY-FIVE STATEMENTS AND SIXTY PUPILS)

Question Numbers	Tally of Number of Mistakes	Total Number of Mistakes
1		11
2		38
3		7
4		10
5		23
6		2
7		8
8		1
9		1
10		25
11		38
12		13
13		35
14		9
15		4
16		16
17		36
18		0
19		0
20		3
21		1
22		15
23		17
24		20
25		24

STEP 3. REARRANGEMENT OF TALLY IN GRAPHICAL FORM

A criticism of the quality of these questions becomes somewhat easier if the range of difficulty in the questions is shown by a rearrangement of the questions in their order of difficulty and if, at the same time, that range is shown in graphical form. The steps by which this may be done are as follows:

1. The questions should be rearranged in the order of difficulty, with the easiest questions (those which had the least number of mistakes) first and the more difficult ques-

tions last. The tally sheet reproduced in Table VI shows that the questions numbered 18 and 19, with no errors, were the easiest, followed by the questions numbered 8, 9, and 21, with one error each, and ending with the questions numbered 2 and 11, with thirty-eight errors each. Upon rearrangement in this order of difficulty the rank becomes as shown in Table VII.

TABLE VII. RANKING OF STATEMENTS ON TRUE-FALSE TEST
AFTER TALLYING

Question Numbers	Number of Errors	Question Numbers	Number of Errors
18	0	12	13
19	0	22	15
8	1	16	16
9	1	23	17
21	1	24	20
6	2	5	23
20	3	25	24
15	4	10	25
3	7	13	35
7	8	17	36
14	9	2	38
4	10	11	38
1	11		

2. These results are now ready to be turned into a graphical table or surface. Cross-sectioned paper like that described in the preceding chapter can be used for this distribution, and a base line should be drawn similar to that previously described. Quarter-inch squares are sufficient in this case, and along the base line the squares should be numbered with the question numbers in their order of difficulty, from left to right, as given from top to bottom in Table VII. Then the squares should be outlined above the base line in proportionate height to the number of mistakes for each of the numbered questions which are given in Table VII. Thus the height of the outline above the base line for Question 7 would be eight squares, and the outline for Question 2 would

be thirty-eight squares. When completed the distribution would appear as in Fig. 20.

Had there been a perfect range of difficulty on this test, the graph would have shown a progression similar to that illus-

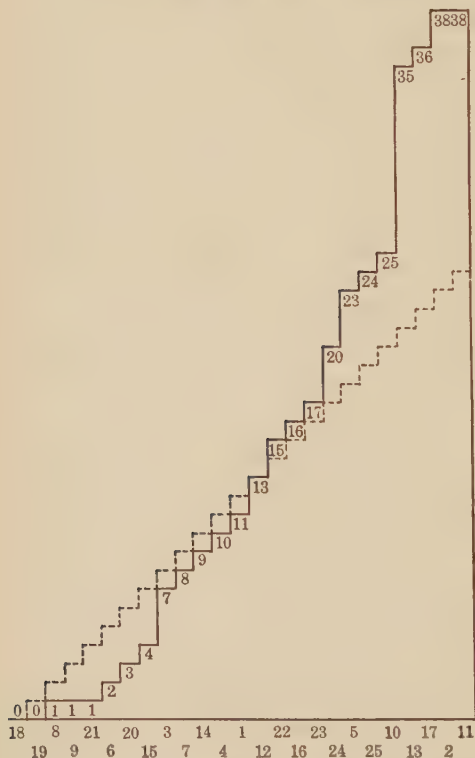


FIG. 20. Graph of question difficulty of True-False Test

Dotted line indicates perfect range of difficulty

trated in Fig. 20 by the dotted line. An analysis of the actual results shows that of the twenty-five questions included in the test five were of almost zero difficulty and that three of these five could have been eliminated without destroying the values of the test. It also shows that there was a very creditable range in the difficulty of the remaining questions, with the exception of Questions 13, 17, 2, and 11, which are very much more difficult than any of the rest. The gap in difficulty between Questions 13, 17, 2, and 11

should be reexamined with a view to determining the difficulties which made them so generally missed. As a matter of fact, in this particular test it was discovered that Question 2 and Question 11 were ambiguous statements to which

an answer of either "Yes" or "No" was correct, depending upon the point of view of the individual who answered, and it was necessary to make a correction in the scoring for both these questions. On the other hand, it was found that Questions 13 and 17 were clearly stated, but that they had been badly taught and that more than half the pupils had failed to make clear distinctions among the principles involved. The interpretation of the questions in the work of diagnosis and the improvement of teaching will be discussed in a later chapter.

2. THE DETERMINATION OF QUESTION DIFFICULTY IN VARIABLY SCORED TESTS

In the case of tests, where the ratings on each element vary from 0 through 1, 2, 3, or more points, such as Completion, Judgment, Selection, and Association Tests, the determination of question difficulty presents a slightly more complicated problem. It is necessary to determine not only the number of times that each question has been given a zero score but also the number of times that each question has received grades other than zero. This may be done as follows:

STEP 1. PREPARATION OF QUESTION-DIFFICULTY TALLY SHEET

The tally sheet for determining the question difficulty on tests of this sort should take a form which will allow separate tabulation to be made on each size of score. In a test, for instance, where there are four possible scores for each question, as is possible in a Completion or a Judgment Test, the tabulation form shown in Table VIII has been used with success. Since a question may be right or wrong or partly right, the problem is to find out the total amount of rightness or wrongness for each question.

TABLE VIII. SAMPLE TALLY SHEET FOR A VARIABLY SCORED TEST

QUESTION NUMBERS	SCORES			
	3 Box	2 Box	1 Box	0 Box
1				
2				
3				
4				

When the test which is being tabulated has only three possible units in the scores, — 0, 1, and 2, — the tally sheet may be modified to contain only these three possibilities. When it contains more than four, the tally sheets should be extended to include the greatest number of possibilities thereby made necessary.

The procedure in using this tally sheet is to record each of the scores that is made and the number of papers that are being tallied. There are two ways of making the tally. One is to tally each question separately, going through all the papers for the scores of Question 1, repeating the entire operation for Question 2, and so continuing with the other questions of the test. This is probably the most accurate method for the teacher who is unaccustomed to the collection of data of this kind, but it is also a time-consuming and laborious method, since it means handling each sheet as many times as there are questions. The second method is to make the tally for the entire set of elements from each sheet before going on to the next sheet. If the teacher is careful to keep in mind not only the scores but also the relative location of the scores with respect to the questions, this method of tallying can be done very quickly and efficiently.

For illustration, suppose that the test is a Judgment Test of ten elements, each of them graded on a scale of four possibilities, from 0 to 3, and that there are thirty-eight

papers included in the test. For the ten questions on the first paper the scores stand as in Table IX.

The teacher should divide these scores into as many groups as necessary, depending on how many scores he can keep in mind accurately while performing the necessary operations for tallying. These operations consist in keeping a number of the scores in mind, such as the first five (3, 0, 2, 3, 0), and at the same time locating each score in its proper question line and in the proper box or score-tally space. At first the teacher will find it difficult to keep more than two or three of these elements in mind, but with practice the number can easily become greater; for the other elements, such as locating the right question line, become practically automatic. It is probably not wise for a teacher to attempt to remem-

ber more than five scores at a time, which in this test would mean that the scoring for each paper would have to be done in two sections: first, the scores on Questions 1, 2, 3, 4, and 5; and secondly, those on Questions 6, 7, 8, 9, and 10.

With the tally sheet in front of him the teacher looks at the scores on the first five questions on the top test paper. In this case these scores are 3, 0, 2, 3, 0. These should be attentively repeated twice and then set down on the tally sheet as follows: A tally stroke is made first in the 3 box opposite Question 1; then in the 0 box opposite Question 2; then in the 2 box opposite Question 3; then in the 3 box opposite Question 4; and finally a stroke is made in the 0 box opposite Question 5.

With this completed the teacher can turn his attention to the second five questions on the top sheet and learn the

TABLE IX. SCORES OF ONE PUPIL IN TEN JUDGMENT-TEST QUESTIONS

Question Numbers	Pupil's Score
1	3
2	0
3	2
4	3
5	0
6	2
7	2
8	3
9	1
10	3

scores for that section. In this case the scores for the five questions are, in order, 2, 2, 3, 1, 3. After repeating these to check the first reading the teacher may turn to the tally sheet and tally the scores by putting a mark, first, in the 2 box for Question 6, in the 2 box for Question 7, in the 3 box for Question 8, in the 1 box for Question 9, and finally in the 3 box for Question 10. With a little practice this can be done both quickly and easily. With the completion of the first paper the tally sheet would have the appearance of Table X.

TABLE X. APPEARANCE OF TALLY SHEET AT CONCLUSION OF TALLY OF FIRST PAPER

QUESTION NUMBERS	SCORES			
	3 Box	2 Box	1 Box	0 Box
1	/			
2				/
3				
4	/			
5				/
6		/		
7		/		
8	/			
9				
10	/			

In doing this work by this method it is possible to make mistakes, especially in locating the right question line. There are probably fewer mistakes made in locating the right boxes; but since one is continually swinging from one box to another as well as from one question to another, one has to be especially attentive with each operation. For this reason it has been found to be a saving of time in re-checking to divide the papers of the class into groups of ten and to check systematically at the end of every ten papers. This checking consists simply in adding together the marks in

each of the boxes for each question, to make sure that each question has the full total of tally marks. If this is not the case, it is necessary merely to re-tally the last ten papers in order to discover the mistake that has been made, which, toward the end of a tally, is much easier and quicker than re-tallying an entire group of papers. In order to know which were the marks that were made on the last ten papers it is necessary for the teacher to place a distinctive mark after the last tally mark made in each box when each ten test papers have been completed, which will determine the tally marks that were added by that particular group of papers. At the end of the checking for the first ten papers tallied the tally sheet will look somewhat like that given in Table XI. The line after certain tally marks indicates the last mark made in that group before going on to the next group of papers, and indicates the closing of that section of the tally. A small cross (X) in a box indicates that no tallies were placed in that box during the tallying of the ten papers. It will be noted in Table XI that if the marks are added horizontally the total for each question considered is ten; this indicates a check on the accuracy of the tally.

TABLE XI. TALLY SHEET AT CONCLUSION OF TALLY OF FIRST TEN PAPERS

QUESTION NUMBERS	SCORES							
	3 Box		2 Box		1 Box		0 Box	
1	////		//		/		///	
2	/		/		////		////	
3	/// //		///		x		x	
4	/		/// /		///		x	
5	/		x		x		/// ///	
6	/// /		///		/		x	
7	/// //		//		/		x	
8	//		/// /		x		//	
9	/		/		//		/// /	
10	///		/		/		///	

If it were found, for example, at the conclusion of the twentieth paper that one of the questions had only nineteen tally marks, it would be an indication that one of the papers had been slighted for that question. Instead of going back over the entire tally process it is merely necessary to go back over the last tallies in each box. As the tallies are checked in the re-checking process, they will appear singly on the different test papers; therefore the teacher should indicate in some way that they have been checked. A simple way to do this is to place a dot *under* the corrected tally, so that when the error is found it will be noticed. In this way a mistake can

TABLE XII. TALLY SHEET AFTER TALLYING THIRTY-EIGHT PAPERS

QUESTION NUMBERS	SCORES			
	3 Box	2 Box	1 Box	0 Box
1	//// / X //// ----- ////	/// / /// // ----- ////	/ / /// / ----- /// /	/// // ----- /// /
2	/ //// /// // ----- ///	/ X X X ----- ////	/// X X X ----- ///	/// /// / ----- ///
3	/// // /// // ----- /// /// ///	/// /// // ----- ///	X X X X ----- ///	X X X X ----- ///
4	/ / X / ----- ///	/// / /// ----- /// ///	/// X X X ----- ///	X /// / ----- ///
5	/ / X X ----- ///	X X // / ----- ///	X /// / ----- /// /	/// /// // ----- ///
6	/// / /// // ----- /// /// ///	/// /// // ----- ///	/ X X X ----- ///	X X X X ----- ///
7	/// // /// // ----- /// /// ///	/// X X / ----- ///	/ / / X ----- ///	X X X X ----- ///
8	/// / /// // ----- ///	/// / /// / ----- /// //	X /// // ----- ///	/// / // ----- ///
9	/ X X X ----- ///	/ X X X ----- ///	// X X X ----- ///	/// / /// /// ----- /// /// /// //
10	/// /// // ----- /// / /// /	/ X / X ----- ///	/ X X / ----- ///	/// /// / ----- ///

be found without the necessity of covering all the previous work, and the tally can be kept correct. The time required to count the tallies and draw a block line after the last tally in each group, or to make the crosses indicative of no tallies in that group, is negligible in view of the time it may save. The teacher can consider it a form of time insurance.

When the entire thirty-eight papers have been tallied, the tally sheet in this case would appear as in Table XII.

TABLE XIII. NUMERICAL TALLY OF THIRTY-EIGHT PAPERS, DERIVED FROM TABLE XII

QUESTION NUMBERS	SCORES			
	3 Box	2 Box	1 Box	0 Box
1	4, 1, 0, 4	2, 1, 2, 2	1, 6, 3, 1	3, 2, 5, 1
2	1, 4, 7, 3	1, 0, 0, 0	4, 0, 0, 0	4, 6, 3, 5
3	7, 7, 8, 5	3, 3, 2, 3	0, 0, 0, 0	0, 0, 0, 0
4	1, 1, 0, 1	6, 3, 7, 3	3, 0, 0, 0	0, 6, 3, 4
5	1, 1, 0, 0	0, 0, 2, 1	0, 5, 6, 5	9, 4, 2, 2
6	6, 8, 9, 7	3, 2, 1, 1	1, 0, 0, 0	0, 0, 0, 0
7	7, 9, 9, 7	2, 0, 0, 1	1, 1, 1, 0	0, 0, 0, 0
8	2, 1, 2, 2	6, 6, 5, 2	0, 2, 2, 2	2, 1, 1, 2
9	1, 0, 0, 0	1, 0, 0, 0	2, 0, 0, 0	6, 10, 10, 8
10	5, 4, 6, 6	1, 0, 1, 0	1, 0, 0, 1	3, 6, 3, 1

The information contained in Table XII, converted into number units instead of tallies, is shown in Table XIII.

STEP 3. CONVERTING TALLY SCORES INTO QUESTION-DIFFICULTY RATINGS AND GRAPHS

When the tally sheet has been completed, the teacher will probably find that the whole distribution will be easier to work with, besides being easier to handle for conversion into a question-difficulty graph, if the numbers of the scores which the tally sheets reveal are transferred to a new sheet of paper in the following fashion. Using the same general

outline as for the original tally sheet, the teacher should make out a fresh sheet and in the various boxes insert the final totals of the scores. This sheet, for the case cited above, is as shown in Table XIV.

One way of revealing a certain amount of truth which is contained in these figures is to make a graph of each separate question, which will tend to show more clearly the differences

TABLE XIV. FINAL TOTALS
AFTER COMPLETION OF TALLY

QUESTION NUMBERS	SCORES			
	3	2	1	0
1	9	7	11	11
2	15	1	4	18
3	27	11	0	0
4	3	19	3	13
5	2	3	16	17
6	30	7	1	0
7	32	3	3	0
8	7	19	6	6
9	1	1	2	34
10	21	2	2	13

between the questions. It will not, however, show much that is of value and cannot reveal the degrees of difficulty between the various questions, because, as the scores stand, it is impossible to give any approximate ranking in question difficulty.

A way that has been found valuable for accomplishing this is to convert these variable scores into a single score for each question. In the case of the True-False Test previously cited it will be noticed that the difficulty was determined by finding the number of instances

in which each element was found to be incorrect. The same method will bring about the same result here, but with a little more complicated calculation. The number of scores that fall into the 0 Box represent answers that are totally incorrect; but the number which fall into the 1 Box are only two-thirds incorrect, and the number which fall into the 2 Box represent questions only one-third incorrect. If, therefore, the teacher finds the sum of the scores in the 2 Box multiplied by one third, and the sum of the scores in the 1 Box multiplied by two thirds, and the sum of the scores in the 0 Box as they stand, the result will be analogous to that used in the True-False Test, a number representing the degree

of difficulty of each question. On Question 1, for instance, it will be found that seven pupils got the question one-third wrong, eleven pupils got it two-thirds wrong, and eleven pupils got it three-thirds wrong. The degree of difficulty is found by adding one third of 7 ($2\frac{1}{3}$) to two thirds of 11 ($7\frac{1}{3}$) and that to three thirds of 11 (11), giving a total difficulty of 21. For Question 2 the result is found by adding one third of 1 ($\frac{1}{3}$) to two thirds of 4 ($2\frac{2}{3}$) to three thirds of 18 (18), which gives a sum of 21. Table XV shows the work in detail for the ten questions, the results being reduced to the nearest whole numbers; the data were taken from Table XIV.

TABLE XV. CALCULATIONS FOR DETERMINING DEGREE OF QUESTION DIFFICULTY OF STATEMENTS IN TABLE XIV

Questions	0/3 3 Box	1/3 2 Box	2/3 1 Box	3/3 0 Box	Total
1	0	$2\frac{1}{3}$	$7\frac{1}{3}$	11	21
2	0	$\frac{1}{3}$	$2\frac{2}{3}$	18	21
3	0	$3\frac{2}{3}$	0	0	4
4	0	$6\frac{1}{3}$	2	13	21
5	0	1	$10\frac{2}{3}$	17	29
6	0	$2\frac{1}{3}$	$\frac{2}{3}$	0	3
7	0	1	2	0	3
8	0	$6\frac{1}{3}$	4	6	16
9	0	$\frac{1}{3}$	$1\frac{1}{3}$	34	36
10	0	$\frac{2}{3}$	$1\frac{1}{3}$	13	15

If a rearrangement of these statements is made in the order of their difficulty, from the easiest to the most difficult, the retabulated order is as shown in Table XVI.

It is easy to see that there is in this short test a very wide range in the difficulty of the questions, as the difficulty ranges from a question that has a difficulty score of 3, out of a possible 38, to a question that has a difficulty score of 36, out of a possible 38. This would have been a better test, as the distribution for the entire class in original scores would probably show, if there had been only one question of the difficulty of Questions 6, 7, and 3, if between them and the

question of the difficulty of Question 10 there had been included two or three other questions, if there had been only one question of the difficulty of Questions 1, 2, and 4, and if between Questions 8 and 5 there had been two or three others.

Question 5 and Question 9 should be examined to determine the general difficulties which there prevailed; for it is evident that those two questions present difficulties beyond

the ability range of the class.

TABLE XVI. REARRANGEMENT
OF TABLE XV IN ORDER OF
QUESTION DIFFICULTY

Question	Difficulty
6	3
7	3
3	4
10	15
8	16
1	21
2	21
4	21
5	29
9	36

It would perhaps be found that these questions presented either ambiguities or unfairness of some sort and should have been eliminated, or else that the ideas which they represent had been badly or inadequately taught.

The range of difficulty becomes more evident if it is assembled in graphical form, as shown in Fig. 21. Here the gap in difficulty between Question 3 and Question 10 is very

apparent, as is also the great gap which intervenes between Question 4 and Questions 5 and 9.

The dotted line in Fig. 21 shows how the question difficulty would have appeared, had there been a more equitable distribution of difficulty.

3. THE DETERMINATION OF QUESTION DIFFICULTY IN BATTERIES OF TESTS

When a battery of tests is combined into one examination, it is of interest and frequently of value not only to determine the relative difficulty of each question in the tests as compared with other questions of the same part but also to construct for the whole test a single table which will show the

relation of any question in the entire examination to every other question. The procedure for accomplishing this is as follows:

STEP 1. FINDING QUESTION DIFFICULTY OF SEPARATE PARTS

The first step is to determine, as has been shown above, the question difficulty for each part of the test separately. If, for instance, there are three parts to the test, — a True-False Test of twenty elements, a Completion Test of seven elements, and a Selection Test of ten elements, — the question difficulty of the three tests should be calculated for each test separately. The difficulty of the questions in the True-False Test should be calculated by taking the number of errors made in each question. The difficulty on the Completion Test should be obtained by taking the sum of one third of the errors in the 2 Box, two thirds of the errors in the 1 Box, and the total of the errors in the 0 Box. The difficulty on the Selection Test should be calculated by taking one half of the errors in the 1 Box and adding to that the total of the 0 Box, when there are only three possibilities in the grading: 0, 1, and 2. If the correcting is of only two possibilities, 2 and 0, then all that is necessary is to take the number of errors in the 0 Box for the degree of question difficulty.

When two or more parts of a test are calculated separately,

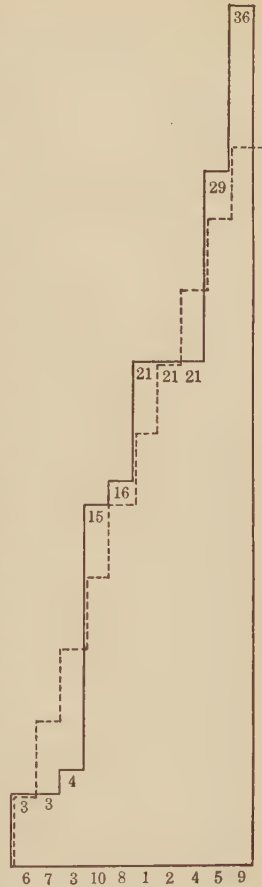


FIG. 21. Graph of question difficulty of Judgment Test

Dotted line indicates perfect range of difficulty

it is neither wise nor desirable to make a graph of each separate part; this can be done later if it is found necessary.

STEP 2. COMBINING THE VARIOUS PORTIONS OF THE TEST

When the question difficulty for each part of the test has been determined, the numbers which represent the relative difficulty of all the questions are ready to be combined into

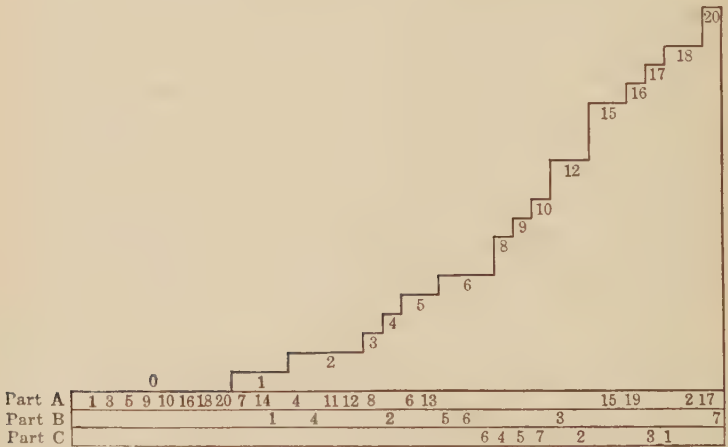


FIG. 22. Graph of question difficulty of a battery of tests

a single scale for the entire examination. The process for the entire group is the same as it would be for any part of the test. The questions should be renumbered, however, by giving them a letter as well as a number designation. The letter should indicate the portion of the test of which the element is a part, and the number should indicate the particular position which it holds within its respective test. An examination of three parts might have the first part, say True-False, labeled A; the second part, Judgment, labeled B; and the third part, say Selection, labeled C. Then an element labeled A-4 would be the fourth element of the

True-False part, B-6 would be the sixth Judgment selection, and C-10 would indicate the tenth Selection unit.

The questions may then be ranked from easiest to hardest, according to the relative difficulty as represented by the difficulty values found for them. When this has been completed the entire group can be made into a graph.

Fig. 22 shows a graph made from the battery of tests given on pages 188-190, consisting of seven Judgment units, twenty True-False elements, and seven Completion sentences, with the proper designations as above described.

Chapter summary. The method for the determination of question difficulty is designed to bring out the number of pupil errors, or the extent of pupil error, made in Teacher's Classroom Tests, with a view to finding out, first, the range of pupil error and, secondly, the causes underlying the commission of the errors. It is a simple method, consisting of a number of clearly defined steps, and it can become almost automatic in its operation. The results which are secured form a solid basis for educational diagnosis as well as for indicating those places where teaching can be improved, and as such they are well worth the effort and the time which they cost.

CHAPTER XIII

SPECIFIC USES: THE USE OF TESTS FOR EDUCATIONAL DIAGNOSIS AND THE IMPROVEMENT OF TEACHING

Educational diagnosis involves working with pupils. One of the important uses of Teacher's Classroom Tests lies within the boundaries of the classroom itself and offers to the teacher a valuable means, not only for determining future work in the classroom but also for evaluating the work that has been done. It shows the point which has been reached by the pupils as well as the point from which further efforts should start. It not only indicates the point from which the work should start but also in some measure indicates the road which should be followed. One phase of this use lies with the pupils themselves. An analysis of the test results will show the teacher where the pupils have failed in their efforts, and at the same time will indicate which particular pupils have failed and upon what particular phases of the work the failure was dependent. These points should be clear to the teacher, though the extent to which advantage is taken of the information lies with the teacher himself. His next work is to use these definite results in taking the further step which is necessary wherever it is possible, namely, determining the causes which were responsible for the undesirable results, and eventually, through that, taking the necessary steps to correct the difficulties. This work, the success of which lies largely in the degree of insight and resourcefulness which the teacher can exhibit, is known as educational diagnosis. It has much the same relation to the teacher as medical diagnosis has to the physician. A teacher may, even with a knowledge of the elementary-school curriculum, with his tests, and with their results, fail to

diagnose the difficulties that are present and may fail to prescribe for the educational ills of his pupils. Such a teacher can be classed with an incompetent physician who, with his stethoscope, his thermometer, and his knowledge, fails to make use of his findings in the constructive remedial treatment of his patients.

The equipment of the teacher must include not only a knowledge of what to teach and why to teach it, but also a knowledge of how to teach. How to teach is not alone a result of experience, since children differ from one another, but it is also a matter of educational diagnosis; and educational diagnosis itself can come only as a result of a measurement of what has been taught. For the correction of the educational ills of pupils, diagnosis is as necessary as experience, and both must be used in the application of the remedial measures, the diagnosis showing where the measures should be applied and the experience of the teacher showing what measures may be applied.

Improvement of teaching involves the teacher. A second phase of the use of tests lies not so much with the pupil as with the teacher and serves as the means whereby the teacher can accumulate the experience which is necessary to follow up diagnosed results. This phase may be called the improvement of teaching, and it is fully as important and just as necessary as educational diagnosis. The problem of self-analysis and self-improvement is very difficult unless some objective standard from which judgments of actual teaching can be made can be pointed out either to a teacher or by a teacher. This has always been the problem and the difficulty in much of the traditional supervision, — the lack of objective standards by which to measure the success of teaching. Teacher's Classroom Tests will not solve the whole problem, but they will aid a teacher in a self-analysis that is both penetrating and instructive, and the use of Standard Tests in fields where they have been perfected will help the teacher much further along in the process. When a teacher once dis-

covers that through the proper interpretation of test results he can make constructive criticisms of his own work, criticisms which, if followed out with remedial measures, will result not only in more efficient but also more effective teaching, he will be more likely to seek such criticisms than to avoid them.

The determination of the pupils most in need of diagnostic attention. Educational diagnosis begins with the steps that have been outlined in the preceding chapters. When the Teacher's Classroom Tests have been given and the papers have been scored, diagnosis can begin. The first step consists in the calculation of the class errors and the determination of the question difficulty, as has been outlined in Chapter XII. This determination will show the characteristic errors of the class, but it will not show the causes of the individual errors which the pupils have made. These can be determined only by an analysis of the papers of the individual pupils; yet, if this analysis were carried out for all of them, though of great value, it would be a long and laborious process and one which in the limited time available to most teachers would be difficult to complete. The teacher must decide for himself just what pupils most need diagnostic attention and concentrate his efforts on them, just as most physicians must concentrate upon patients who are actually ill.

It is indeed difficult for the teacher to determine which of all his pupils need some diagnostic work, but it is not difficult for him to decide which ones need it most. They are either the pupils who cluster about the foot of the class (who cluster, as far as the tests are concerned, about the low end of the frequency distribution of the scores) or those who, as is shown in a later chapter, are far below their achievement possibilities. By consulting the frequency surface the teacher should determine how many of his pupils he plans to analyze carefully from a diagnostic point of view. It is wise for the teacher to start with only a few, not more than two or three, until the success of his efforts with those pupils enables him to know that his methods will produce results.

Question analysis for diagnostic purposes. The papers written by the two or three pupils who have accomplished least should be singled out from the entire pile of papers and analyzed carefully. The selection of these papers can be determined by consulting the frequency surface and picking out the papers whose scores are there written down. From these papers, individually, the teacher should extract the questions which are answered incorrectly but were not so answered by the class in general. They will probably give the surest indication of the underlying causes responsible for the difficulties of the pupil. These questions can be determined by a comparison between the pupil's scores and the table or graph of question difficulty which has been constructed. The questions which a large proportion of the class have answered incorrectly (the questions with the highest difficulty scores) are not likely to show any difficulty which is characteristic of this lower group. At this time, therefore, these questions can be ignored, but by comparison with the question difficulty those questions which were answered correctly by the majority of the class but answered incorrectly by these pupils of lowest scores can be quickly determined.

In the two accompanying illustrations are given a frequency surface (Fig. 23) and a question-difficulty graph (Fig. 24) of the same test. The two pupils whose scores are shaded in the frequency surface are those whose work is to be diagnosed. In the question-difficulty graph the characteristic errors of this group are probably to be found in the questions which have been easiest for most of the class.¹ These have been shaded in Fig. 24. Questions of greater difficulty are characteristic of the class as a whole. That which these two pupils were unable to do and which the large majority of the class was able to do constitutes the question at issue. This is represented by Questions A-7, A-14, A-4, A-11, A-12,

¹ There is no arbitrary standard for this determination. The writer, in his own field, has found that a determination of from six to eight questions has been conducive to good results.

B-1, and B-4. Which of these questions was missed by each of these pupils? By going over the two test papers and by watching these questions only, the particular elements which for most pupils were easy but for these pupils were hard can be found.

Determination of causes of difficulties. The next step is for the teacher to make some determination of the causes that impelled these pupils to make the errors. Here all the teacher's knowledge of the pupils, all his knowledge of teaching, and all his insight and resourcefulness must come into play, because the results may have any of an almost infinite

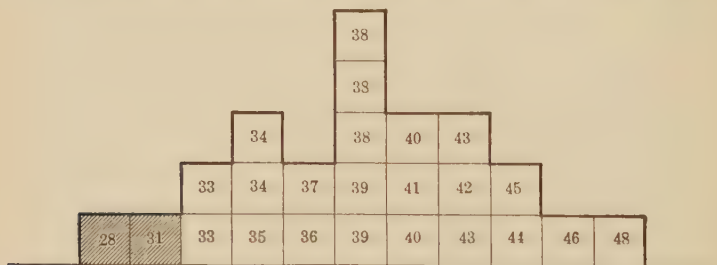


FIG. 23. Frequency surface of class

Shaded portion indicates pupils to be diagnosed

variety of causes. The teacher's work is to find the most likely among these causes, to concentrate upon them, to devise remedial measures for them, and then, at a later time, to make such tests as will indicate whether the remedial measures have been successful. Because of the infinite variety of causes and because of the impossibility of anticipating the causes which teachers find in their work, all that this chapter can do is to outline for the teacher a course of investigation which will serve to indicate, at least, the likely causes of the failure of pupils.

1. *Physical causes.* How far physical causes can interfere with the work of an individual depends somewhat upon the way in which a test has been given and as well upon the

requirements which the test makes of the pupils from a physical standpoint. If a test has been dictated, for example, deafness might be a real cause of misunderstanding and a source of resultant error, whereas defective eyesight might be only a minor possibility save in the writing of the answers. On the other hand, if a test has been given by the blackboard

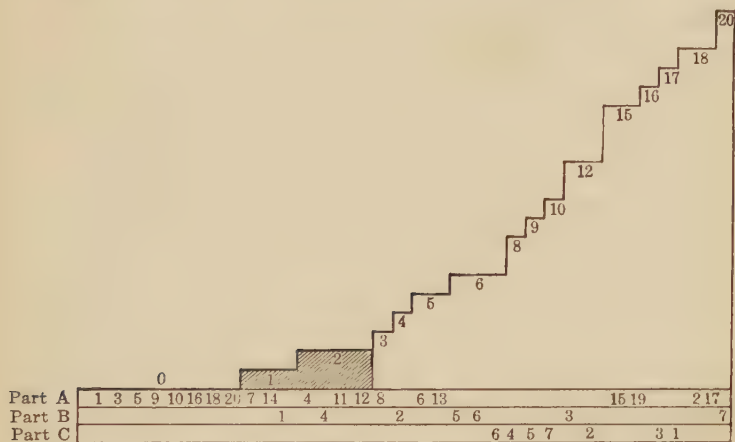


FIG. 24. Question-difficulty graph

Shaded portion indicates questions to be used in diagnosis

method, deafness might be a minor factor and defective eyesight a major factor as a real source of misunderstanding. Physical defects, ranging from mild or temporary difficulties, such as headaches and sleepiness caused by bad ventilation, to severe and chronic troubles, such as defective eyesight and malnutrition or communicable diseases, are commonly prevalent among school children, and the more severe forms are even more prevalent, probably, than most teachers suppose. In the writer's recent experience, for example, a teacher who was having difficulty with a class came to the conclusion that the pupils were not seeing properly and had the entire class examined individually by a competent and interested oculist. The results were astounding; they showed that fully

60 per cent of the class were in need of corrective measures with respect to eyesight and that of these more than two thirds needed glasses in order to see with any accuracy at all.

Dr. Thomas D. Wood has grouped into three divisions for the use of teachers the signs of health disorders and physical defects in school children, as follows: (1) indications which teachers should be trained to notice and report because they point to disorders that may have serious consequences and require delicate adjustments; (2) signs of abnormality pointing to more chronic disorders, which should be remedied early; (3) indications of disturbance which are important in connection with other signs of physical disorder. This tabulation, as reported in the transactions of the Fourth International Congress on School Hygiene and modified somewhat by Dr. Wood for inclusion here, will be found at the end of this chapter and should be consulted by the teacher as a first step in the diagnosis of physical difficulties of children. Taken in connection with test failures, these disorders assume a new significance, and it is possible that they may be either direct or indirect causal agents.

Should physical defects be found in such proportion as to make it even possible or desirable, Dr. Wood recommends that the pupils be grouped according to these defects in the same manner as pupils of varying levels of mentality might be grouped.

When an examination of the pupils under consideration has been made to determine the extent to which physical defects might be the cause of the difficulties discovered in the tests and no likely causal agent has been discovered, the teacher should prepare for the next step in the analysis.

2. *Mental causes.* The second major step in this analysis has to do with possible mental causes of the failure of the pupils concerned. Therefore the teacher should next question the ability of these pupils to be a part of the class group. The inquiry relates to the extent to which the pupils are inherently capable of doing the work which they have been

asked to do as a result of their placement in the group. If the child has normal ability and no physical defects which would tend to lower the quality of his work, he should be expected to do the work which the other pupils in the class are doing, and the reasons for his failure to do so must be found in some other field than his mentality. On the other hand, if the pupil has not normal ability, he cannot be expected to do the work of children of superior ability and do it at the same speed. Remedial measures with subnormal pupils take a different form from that usual with normal and supernormal children. The subnormal pupils need a special education and a special training, and such children should be reclassified and placed in a group where they can have special attention. In such a group they will be pitted against equals in ability and will be able to move at the same rate as their classmates as well as receive the increased and probably more sympathetic help which they need.¹

The determination of this ability necessitates the use of an accepted and reliable mental, or intelligence, test, which should be administered by an expert. This test will determine, with more accuracy than any other method at present known, the degree of ability with which a pupil is endowed for accomplishing the work that is asked of him. The information which such tests give should be in the hands of every teacher, and every teacher should be taught their use and limitations. Without such a test a teacher is handicapped in further diagnosis, but with it he is equipped for a reasonable and useful determination of the needs of his pupils.

Where the teacher finds it impossible to have a reliable intelligence, or ability, rating made of his pupils, he must accept the limitation imposed, and until such a rating is available assume that the difficulties of his pupils are not based upon mental inability to do the work that is required.

There are undoubtedly certain difficulties influencing

¹ Compare L. S. Hollingworth's *Psychology of Subnormal Children*. The Macmillan Company, New York, 1923.

mental status other than these that have been mentioned, which carry into specialized fields of great importance but which are difficult of diagnosis and hard to adjust. There are mental difficulties caused by many sorts of maladjustment of children. Social maladjustments (ranging from mere grade misplacement to major misplacements in any portion of the social environment), racial maladjustments of all kinds, nationality maladjustments (of which language is a relatively minor factor), and that great field of which so little is positively known, emotional maladjustments, are all intimately related to this problem. With regard to this matter the Report on "Health Education" of the Joint Committee on Health Problems in Education has many good suggestions for teachers, which are quoted in part at the end of this chapter.

When the teacher has made his judgments with respect to the mental character of his pupils, he can go on to the field of possible academic causes in the further diagnosis of the particular pupils under consideration.

3. *Academic causes.* Here, for the first time, the teacher reaches a field of possible causes that is directly concerned with his teaching and where, undoubtedly, in the great majority of cases it may be concluded that the real difficulties encountered by the pupils will be found. These causes may be roughly classified into three divisions, each of which requires a different sort of remedial work, and each of which is of the greatest importance for a teacher to determine and to remedy wherever found.

a. *Mechanical causes.* The first of these three divisions may be thought of as the mechanical causes, or mechanical defects, which might be responsible for the failure of the pupil. One of these mechanical causes may be a lack of understanding of the method of taking these tests. If a pupil does not know what to do in order to indicate what his answer is, his answers as they are read are more likely to be incorrect than they are to be correct. If an inspection of

the test results shows that the pupil has probably made his answers or selections at random and without any demonstrable purpose, it may be concluded that he does not know what he should do. This is the case when the markings put down are not the markings that were asked for, or when the answers made indicate a misinterpretation of what was wanted. A second mechanical cause frequently results from a lack of comprehension in reading. The pupil may be able to read the words without knowing exactly what they mean; in this case he is at a disadvantage in making his answer to the question. This is frequently so in the Judgment Test, when it is found that the answers given are totally aside from the question in hand. If this cause is suspected, the teacher should watch the pupil carefully in other fields for signs of lack of ability in reading; and if this inability is found, it would confirm the diagnosis and make necessary the application of remedial measures in reading. The same may be said for any of the basic skills, such as those of handwriting and spelling. A deficiency in any of these is the forerunner of failure in the tests when given in certain forms and should become the basis for immediate remedial measures or else the reclassification of the pupil.

b. Lacks-in-knowledge causes. The second of these three divisions of academic causes may be thought of as the lacks in knowledge which cause failure. These are of course self-evident in most of the tests, although they are likely to be secondary causes rather than primary ones. If a child's paper indicates that he lacks the knowledge to answer the questions, it is probably true that this lack is caused in its turn by some of the other causes given in this outline. The fact that a pupil does not know a thing which his classmates do know may be merely indicative of a lack of knowledge and nothing more, but it is usually indicative of a lack in ability or a lack of interest or a lack of attention or a lack of physical well-being, or some other specific lack the natural result of which is a lack of knowledge.

c. Lack of proper attitudes. The third of these three divisions of academic causes may be thought of as a lack of proper attitudes. A lack of interest in the work in hand is sure to bring as a result a future lack of specific knowledge of the thing itself. A lack of attention in class or during the progress of the test is of course a large source of error. These lacks are fairly easy for the teacher to see, though the remedies are more difficult to discover. Fatigue, of whatever nature, is a potent factor undoubtedly in determining the attitude of a pupil toward his work. Psychological investigation seems to show that mental fatigue is rather rare among school children, but that physical fatigue, caused by improper school or home conditions, excessive emotion or excitement, and the like may react adversely on school work.

The experience of teachers who have used Teacher's Classroom Tests seems to indicate that the attention of pupils and their interest in school work are much increased through their use of the tests, which present in themselves a remedy for many of these lacks. Pupils who have failed to respond to other methods of arousing interest have attacked the tests with vigor and have been stimulated to classroom effort through the connection of the tests with their work. Undesirable attitudes, slipshod methods of work, inaccurate thinking, and carelessness are revealed to the pupil himself through the results of the tests and are proved to him, so that in many cases the continued use of the tests provides the only remedial work that is needed in this connection.

All the tests lend themselves to this sort of use, and the teacher should not fail to extend that use wherever possible. Perhaps the most potent means of bringing home to a child the mistakes which he has made and the needs for his own improvement consists in the after-test treatment of the results with his class. Each child should of course have his own paper returned to him and should be given the opportunity to find out each of his mistakes. If he knows why he

made the mistakes, so much the better; but if he does not know, he is likely to be anxious for suggestions. Each pupil should also be shown where he stands with relation to the rest of the class. Pupils will quickly learn to read the distribution surface, and it is good policy to copy it on the blackboard, placing in the proper squares the score numbers, so that each pupil can locate his own score and thereby see his relation to the other class members. The use of colored chalk, and also the pointing out such characteristics of the curve as that those who have reached scores above a certain point have done well within the group and that those who received scores below a certain point have done less well than they should, will make the curve have meaning and thus offer a means of stimulation to further or renewed efforts. Probably one of the most potent arguments for a proper classification of pupils in a school is the fact that a child who would continue invariably to be at the bottom of the pile, who would continually fail, in other words, had better be in a lower group where his efforts would be more on a par with those of his fellows. Where pupils are so placed that it is impossible to arrange matters without having the same children invariably fail, the system of Achievement Ratios discussed in a later chapter will be a better scheme for judging the efforts of the class.

Question analysis for the improvement of teaching. The use of tests for the improvement of teaching, as has been stated, is related not so much to the individual pupils of the class as it is to the class group as a whole, and particularly to the work of the teacher as it relates to the class group. In this phase of the use of test results the teacher should concentrate not on the minor difficulties as revealed in certain low-ranking papers but rather upon the major difficulties as revealed in the results of the entire class. The analysis of the question difficulty on a test or a battery of tests furnishes the first clue to the needs of the teacher. The major questions in which there was difficulty should be carefully

analyzed to discover in which of two fields the difficulties can be classified. These two fields are the following :

1. *The questions themselves may form a possible difficulty.* One possible cause of the difficulties may be the questions themselves. Ambiguities, misunderstandings, unfair questions, negative statements, catch questions, or questions of undoubted excessive difficulty may all come under this heading. The only remedial work in this connection lies either in the further experience of the teacher in better gauging the abilities of the pupils or in increased skill in writing a good quality of questions. The more of these tests the teacher uses, and especially the greater the use which the teacher makes of after-test interpretations, the more he will improve in his ability to construct a fair and a well-adapted test.

2. *Ineffective teaching may form a possible difficulty.* A second possible cause of difficulties, however, lies in ineffective teaching itself, and the tests furnish an objective standard whereby a teacher can judge his own limitations and at the same time locate some of his major difficulties. It means, of course, a certain self-analysis and a certain humility on the part of the teacher, but most teachers are anxious to improve their teaching and most of them, as well, are willing to admit, to themselves at least, that their teaching can be improved. Having taken this attitude, the teacher can use the test results to make some valuable observations concerning his own efforts. These observations can be aided if the teacher can make such a division of his equipment as will enable him to examine various parts of his teaching separately. Such an attempt is made in the following pages, where the equipment of a teacher is divided into separate parts for the convenience of a teacher in his self-analysis in the light of test results.

Improvement in teaching-skills. The first element in the improvement of teaching lies in the improvement in teaching-skills. One of these skills is in the presentation of the work to

the pupils. Unskillful introductions to new work, uninteresting presentations of subject matter, too little emphasis upon early important details, and too little sensitiveness to the reactions of the pupils are all a part of this lack of skill and may be revealed in the test results. Another of these teaching-skills is the use of drill. Too little drill to fix important reactions, too much drill with unsatisfactory results, drill upon inconsequential details to the exclusion of important and necessary details, or unnecessary drills are all a part of this form of lack of skill. The test results may reveal the evidences of this lack of skill and furnish a means whereby an improvement in it will serve to improve the teaching itself.

It may be also that the teacher has difficulty in utilizing proper illustrations for the work. The lack of this skill causes misunderstandings of a serious nature in future work. Inadequate illustration, illustrations in which there are too few connections with the previous experiences of the pupils, illustrations or analogies which present phases or ideas that are contrary to the purposes intended, or illustrations too artificial to convey the reality which they should, are all evidences of this lack of skill and can easily be detected in the analysis of the test results. The detection of such a lack followed by sincere efforts at correction will serve to improve teaching in that field.

Other phases of skills which the teacher might well question as a result of the test analysis are such elements of teaching as questioning, repetition, meeting children upon their own level, organization of reviews, providing for sufficient amounts of recall, and elements of like nature. Many phases of teaching-skills, or the lack of them, such as these mentioned may be located as a result of the test analysis and may then become a definite goal in the improvement of teaching.

Improvement in teaching-knowledge. A second phase of the improvement of teaching may be classed as the improve-

ment which may result from increased teaching-knowledge. This has at least two large component parts: first, the knowledge of the curriculum and its allied units; secondly, the knowledge of teaching as a science with its allied units.

From the standpoint of the curriculum the tests will measure only the extent to which the teacher has succeeded in teaching the phases of the curriculum which he tried to teach. An analysis of the question difficulty and the major types of question which the pupils in general failed to answer correctly will show very definitely where the teacher failed to teach the elements of the curriculum which he tried to teach, and these deficiencies can thereby be emphasized by the teacher until they are eliminated for that class. For future classes, if the teacher takes advantage of his previous experience the same deficiencies can be eliminated as they occur. Because the teacher himself constructs the test papers, the test results cannot measure the extent to which the curriculum has actually been taught. This is a field for the use of Standard Tests, and in this field the teacher should employ these tests to determine these questions of status.

Other phases than the subject matter of the curriculum, however, may be the underlying reason for the difficulties of the pupils. It may be that the difficulties of the teaching lie in the fact that the teacher has acquired too narrow a range of knowledge of the subject matter; in that case he should make efforts to extend that knowledge in order to be able to teach his pupils with a greater ease and understanding. It may be also that the particular method which the teacher has used with the pupils might be improved. This implies the need of a greater knowledge of possible teaching-methods and a greater insight into their possible uses, although these phases of teaching may not be readily apparent from a simple analysis of the test results.

How much knowledge of teaching as a science a teacher holds may be frequently observed in the test results. This

knowledge embraces all that portion of a teacher's knowledge which may be included in such terms as his philosophy of education, his knowledge of educational psychology, and his acquaintance with the sociological problems of education. They are not mere abstract and theoretical parts of teaching-knowledge placed in a teacher-training curriculum for the purpose of filling in certain spaces between units of subject matter, but they are useful and widely adaptable tools which, too often, a teacher neglects in the exercise of his daily routine. Teachers who do not provide sufficiently in their teaching, for instance, for the Laws of Learning or for the laws responsible for the formation of habits, or who do not understand the nature of children, are neglecting some of the more fundamental elements of the psychological equipment which a teacher ought to have for continual use. Again, teachers who follow blindly the curriculum as laid down, without showing due regard for the purposes of the various phases of instruction, who do not keep in mind the more fundamental reasons for their teaching at all and who do not feel the importance of a definite goal farther off than the printed word of the textbook or the published curriculum, are neglecting some of the useful tools which are provided by an educational philosophy. Further, the teacher in rural schools, acquainted with the curriculum and practice in urban schools, who does not make such adaptations in his work as would take advantage of the wide differences between urban and rural life, is neglecting some of the more fundamental tools which are provided in a knowledge of the sociological foundations of education. These are of tremendous importance in the daily work of the teacher, and certain deficiencies in them are reflected in the test analysis which a teacher can make; and definite efforts to correct the deficiencies will almost inevitably result in improved teaching.

Improvement in teaching-attitudes. The third phase of the improvement of the teacher is concerned with his attitudes

or teaching ideals. These are reflected somewhat in the tests which the teacher makes and gives, and are reflected more in the judgments with which a teacher makes his ratings. It is difficult for a teacher to make an analysis of his attitudes which make for difficulties in learning by his pupils, since these attitudes are usually subtle and connected with much of the philosophy which the teacher holds. If the teacher can, however, point to some evidence in his work of an attitude on his part which makes learning difficult for his pupils, such as an attitude of overemphasis on the subject matter of instruction, which is perhaps the most common result of the greater ease of using subject matter for testing purposes rather than the qualities which it is hoped may be developed from it or the experiences which should grow out of it, and if, in addition, the teacher can make definite and constructive efforts to correct the attitude, improved teaching will undoubtedly result.

Of the attitudes which a teacher may hold with respect to his teaching and with respect to his pupils the three following will be found to reveal themselves most easily in the test results. First, the teacher may regard teaching as a matter of filling a child with a certain amount of subject matter, much as a quart cup might be filled with liquid, rather than as contributing to the growth of his pupils. The overemphasis upon subject matter, and especially upon details of little consequence, is an evidence of this. A second attitude is that of seeing the materials of teaching in the light of some remote and future use rather than as the stuff of most complete present living for his pupils. The motivation with which teachers sometimes try to arouse children by saying, "You should work harder on this, for you will need it when you are grown up," is an evidence of this sort of attitude. The contrary should be the case: to try to make the thing fit into the present needs, the present growth, and the present life of the child. A third attitude consists in considering as the essence of education the facts and information and principles

laid down in the curriculum rather than the experiences which should result from the addition of information and facts and principles. The correction of such attitudes would go far to improve teaching, because they tend to govern practically every act of the teacher.

Chapter summary. Educational diagnosis is concerned with the characteristic difficulties of a few individuals within a class, whereas improvement of teaching is concerned with the characteristic difficulties of a class as a whole. The one is an endeavor to improve the work of certain pupils; the other is an endeavor to improve the work of all pupils through better teaching.

Diagnosis may have physical, mental, or academic aspects, and difficulties of individual pupils may arise from any of these as causes or from any combination of them. In order to discover what these causes may be, the teacher should analyze the errors of the few selected pupils in an orderly way with as full a knowledge of the physical, mental, and academic characteristics as is possible for him to acquire. Out of this knowledge the teacher must devise ways and means of remedying the defects which are discovered.

Improvement of teaching is concerned with the development of improved teaching-skills, the acquisition of greater teaching-knowledge, or the correction of certain teaching-attitudes. Here again the teacher must analyze his test results in an orderly and progressive way, so as to discover where his major difficulties as a teacher may lie and so as to be able to devise some appropriate means of correction.

Because physical and emotional difficulties are frequent among school children, and are potent factors in preventing pupils from achieving as much as they might in terms of their abilities, the following quotations are made. As these quotations indicate, there is much that a teacher can and should do to correct the difficulties among her own pupils. Teachers are urged to read and to follow the suggestions given, in order that improved teaching may result.

SIGNS OF HEALTH DISORDERS AND PHYSICAL DEFECTS
IN SCHOOL CHILDREN

[Arranged for Teachers by Thomas D. Wood, M.D., Teachers College, Columbia University, New York City.¹]

The following signs of disorder have been arranged in three groups for the use of teachers in detecting possible health and physical defects in children under their care.

Group I contains signs of disorder which teachers should be trained to notice and report to constituted authorities.

Group II names signs of abnormality pointing to more chronic disorders which should be remedied early.

Group III contains indications of disturbance which are important in connection with other signs of physical disorder.

GROUP I. *Indications of Health Disorders in Children which Teachers should be trained to notice and to report to Constituted Authorities*

Signs

Nausea or vomiting

Chill, convulsions (fits)

Dizziness, faintness, or unusual pallor (alarming paleness of the face)

Eruption (rash) of any kind

Fever

Running nose

Red or running eyes

Sore or inflamed throat

Acutely swollen glands

New cough

Any distinct or disturbing change from usual appearance or conduct of child

The foregoing signs should be used by teachers as a basis for excluding pupils from school for the day or until the signs have disappeared or until the proper health officer has authorized the return of the pupil to school.

¹ This table is a revision of the original table as given by Dr. Thomas D. Wood in Transactions of Fourth International Congress on School Hygiene, Vol. IV, pp. 135-692.

GROUP II. *Signs of Abnormality pointing to more Chronic Disorders
which should be remedied early*

Signs

Mouth-breathing	}	Disorders of nose, throat, ear, and organs of respiration
Loud breathing		
Nasal voice		
Catarrh		
Frequent colds		
Offensive breath		
Chronic cough		
Deafness		
Twitching of lips		
Headache		

Headache	}	Eye disorders and defects
Crossed eye		
Squinting		
Holding book too near face		

Decayed teeth	}	Teeth defects
Crooked teeth		
Discoloration of teeth		
Offensive breath		

Inability to hold objects well	}	Nervous disorders
Spasmodic movements		
Twitching of eye, face, or any part of body		
Nail-biting		
Perverted tastes		
Sex disturbances		

Pain in feet	}	Defects of feet
Toeing markedly out		
Flat-foot gait		
Swelling, puffiness of feet		
Excessive perspiration of feet		

Unequal height of shoulders	}	Incorrect posture
Flat chest		
Round neck and shoulders		
Stooping		

GROUP III. *Indications of Disturbance which are Important in Connection with Other Signs of Physical Disorder*

Signs

Deficient weight	}	Nutritional and general disorders
Pallor		
Lassitude		
Perverted tastes (food)		
Slow mentality		
Peculiar or faulty postures		
Underdevelopment		
Excessive fat		
Low endurance		
Disinclination to play		
Fatigue		

Pigeon-toed gait	}	Defects of feet and legs, and defective movements
Shuffling, inelastic walk		
Exaggerated knee action in walking		
Shifting from foot to foot		
Standing on outer edge of feet		
Standing on inner side of feet, heels turned out		
Locking knee		
Leaning against wall or desk		
Shoes run over at either side		
Wearing out soles asymetrically		
Twitching of foot muscles		

APPLICATIONS OF MENTAL HYGIENE IN SCHOOL¹

There are important applications of mental hygiene which should be made to the school. It would be desirable to have a complete examination of every school child upon school entrance, this examination to include the child's mental as well as his physical health. This is a goal that is far from being realized, but

¹ Quoted in full from chapter on "Mental Hygiene" (pp. 62-64), from *Health Education*, the Report of the Joint Committee on Health Problems in Education of the National Education Association and the American Medical Association, prepared under the direction of Thomas D. Wood, M. D., chairman. New York City, 1924.

there are still many things which teachers can do. A few suggestions are given :

1. Teachers should help their pupils to acquire emotional control, and should avoid any course of action which will arouse undesirable emotions. Children should never be frightened; a childhood fright may become the basis for an adult psychosis. Children should not be ridiculed, shamed, or embarrassed; a child's fear of ridicule may be so intense as to paralyze effort. There should be a calm, orderly atmosphere in the schoolroom which avoids both undue restraint and emotional excitement.

2. Help the shy, easily embarrassed child to overcome his bashfulness and emotional disturbance, so that he may carry on his work and play with other people more happily and efficiently.

3. Teachers should help their pupils to establish habits of intellectual honesty; to meet problems squarely and not to dodge the issue.

Children should not be lied to concerning important matters, especially about the matter of sex. The lying and deceit are soon discovered, and the experience is exceedingly bad for the child. Much of the unhappiness, worry, and failure at school, and the nervous illnesses of young adolescents, as well as the nervous and mental breakdowns of later life, are due to the misunderstanding of these matters that has been brought about by the lying and deceit of others. It is of very great importance that this be avoided. The questions of a child along these lines should be answered honestly and without embarrassment in accordance with the ability of the child to understand. — FRANKWOOD E. WILLIAMS, *Mental Hygiene and Childhood*

4. The habit of concentrating on the present task is one which should be encouraged. Teachers should help their pupils to learn how to work successfully and efficiently. A certain amount of physical and mental work is healthful. Much unhappiness and mental distress come both to children and adults, from inability to work successfully.

5. Children should be encouraged to find a real solution to each problem that faces them, to meet their problems by activity instead of daydreaming. The daydreaming is not harmful if it issues in activity, but excessive daydreaming which leads nowhere is undesirable.

6. The teacher should make every effort to keep the child from developing a feeling of inferiority. Every child should

have a chance to succeed at something; constant failure establishes the habit of failing, and an almost insurmountable obstacle of discouragement or indifference. Teachers should adjudge success upon a basis of effort and improvement as well as natural ability and achievement.

7. Encourage activities which inherently emphasize the desirable qualities, e.g., coöperative sports, school papers, student government, civic activities, hobbies, development of special talents and abilities, scouting activities.

8. Encourage socially useful activities, and the development of interest in other people's welfare.

9. The adolescent age is characterized by a combination of emotional instability and increasing independence which often results in what appears to be perfectly unreasonable behavior. It is worth the teacher's while to attempt to understand all such occasions, and herself to be not only reasonable, but intelligently constructive in dealing with her pupils at such times.

To sum up, habits of truthfulness and honesty, cheerfulness, unselfishness, helpfulness, sociability, courage, persistence, and resourcefulness should be among those most emphasized.

SELECTED BIBLIOGRAPHY

- WOODY, C. "Informal Tests as a Means for the Improvement of Instruction," pp. 87-94 of *First Year Book*, Department of Elementary School Principals. National Education Association, 1922.
- RUSSELL, C. *Improvement of the City Elementary School Teacher in Service* (Contributions to Education No. 128), pp. 89-103. Teachers College Bureau of Publications, New York, 1922.
- PINTNER, R. *Intelligence Testing* (chap. xi, "The School Child"). Henry Holt and Company, New York, 1923.
- HOLLINGWORTH, L. S. *The Psychology of Subnormal Children*. The Macmillan Company, New York, 1923.
- MCCALL, W. A. *How to Measure in Education* (chap. iii, "Measurement in Diagnosis"). The Macmillan Company, New York, 1923.
- WOOD, DR. THOMAS D. *Health Education*. Committee Report of Joint Committee of the National Education Association and the American Medical Association, 1924.
- COURTIS, S. A. "Educational Diagnosis," in *Journal of Educational Administration and Supervision*, Vol. I, pp. 89-116 (February, 1915).

CHAPTER XIV

SPECIFIC USES: THE MAKING OF COMPOSITE TEST SCORES

The desirability of composite scores. It is frequently desirable for a teacher to combine the results of several tests. It will be found that with the same group of pupils the relative standing of any one pupil will vary from one test to another and that, although for any one test the relative ranking of the pupils is quite clear from the frequency surface, successive tests change the shape of that surface as well as the relative position of the pupils within it. In order to get a semester ranking on a series of tests for a group of pupils in a single subject, or to get a semester ranking in all the tests of all the subjects in which Teacher's Classroom Tests are used, it is necessary for a teacher to combine the results of the tests into a composite score.

A composite score, however, which is made up by merely taking the sum of the various test scores that were made by the several pupils is unfair. As has been previously stated, the difficulty of the tests is variable. One test may be easy while another is difficult, and the scores on the easy test may be twice as large as the scores on the difficult test. In simply adding the final raw scores the easy test would have twice as much weight as the difficult test. There are differences as well in the judgment of the teacher, which would make differences in the total of the test scores from one test to the next. In one test the scorer might use a discrimination of 2 points, and in another the discrimination might be only 1 point. In such a case the total scores on the one test might be twice as great as on the other, with the consequent unequal weight if the scores were simply added for a composite.

In the following illustration there are two frequency surfaces shown, involving the same sixteen pupils. In the first

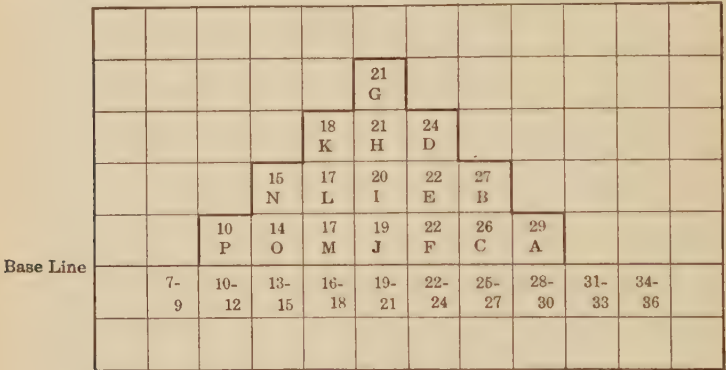


FIG. 25. Frequency surface for sixteen pupils in Test I

test the low score is 10 and the high score is 29, whereas in the second test the low score is 10 and the high score is 57.

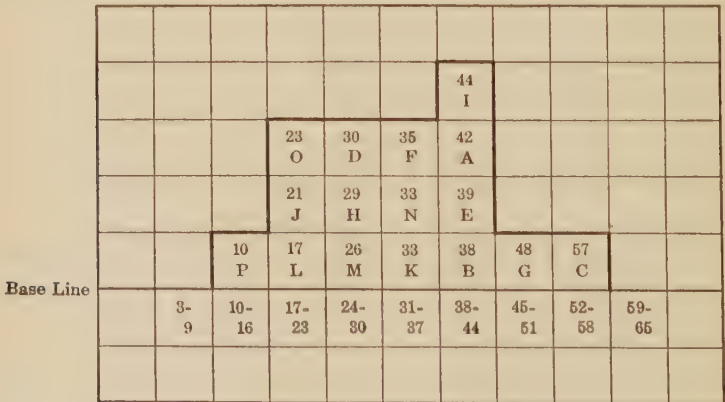


FIG. 26. Frequency surface for same sixteen pupils in Test II

Because the two tests involve the same pupils it is evident that the two test curves represent about the same range of

difficulty, but that Test II has a larger number of possible units of scores for its range than has Test I. Either the discrimination on Test II is somewhat finer than on Test I or there are more elements contained in it. If each pupil's raw scores on these tests were added, pupils who received high scores in the second test would have a distinct advantage over those who received low scores, because for equivalent abilities the raw scores mount more rapidly in the second test, and both tests start with the same low score.

A method should be used which will eliminate these difficulties as far as possible, if a reliable composite score is wanted. This method should make the scores in the two tests comparable, either by reducing both sets of scores to the same scale or else by reducing them to the same kind of unit so that they can be compared on the same scale.

A method for reaching reliable composite scores. A method has been devised whereby the difficulty of the tests can be equated and whereby differences in the test range can be eliminated, and this method makes it possible for a teacher to reduce the scores of the individuals in any one class to the same units, either for comparison or for making composite scores. This method consists merely in reducing the scores of each pupil to a scale score which indicates the pupil's position upon the base line of the frequency surface of each test. The method was devised and used by Dr. William A. McCall¹ in his construction of "T scales" and is fully described in "Scaling the Test," chaps. ix-x, in his *How to Measure in Education*. The method as used by Dr. McCall has been somewhat modified for the purposes of these scores; so the results secured by the teacher with Teacher's Classroom Tests are not T-scale scores, though in appearance and derivation they somewhat resemble them. *The teacher must remember that the scores here obtained are comparable only with other scores obtained in Teacher's Classroom Tests with*

¹ *How to Measure in Education*, pp. 249-306. The Macmillan Company, New York, 1923.

the SAME group of pupils. They cannot be used for comparison with similar scores obtained with other groups of children, even when the same tests are used. The T-scale standard is based on the ability of twelve-year-old children, whereas this scale standard is based upon the ability of the group with which it is used. To distinguish this type of scale from the T scale used by Dr. McCall, it is here proposed to call it an "M scale." The sections which follow show how to find the raw scores from Teacher's Classroom Tests, and the succeeding chapters show how these M-scale scores may be used.

STEP 1. PLACEMENT OF SCORES IN RANK ORDER

The first step in reducing the raw scores to M-scale scores which can be compared directly with other M-scale scores obtained from the same group of pupils is to place the scores in rank order. This merely means listing the scores from greatest to least, or from least to greatest, in the order of their size. It is most convenient to do this directly from the frequency surface, when that has been previously constructed; and it can be done in that way when the actual raw scores have been written in the appropriate squares, as was advised in an earlier chapter. The teacher should begin with the highest scores on the frequency surface and should place them in the order of their size from top to bottom on a sheet of paper, with all the scores (even if there are several scores of the same size) inserted in their proper places.

In the two tests illustrated in Figs. 25 and 26 the pupils may be identified in the two tests by the inserted letters. It is unnecessary for the teacher to insert these letters, because he will have the identifications of the pupils on the original test papers with the original scores. It is impossible to reproduce the original test papers here; so this device is substituted, since the scores must later be identified.

Beginning with the highest scores on each test, the rank order of scores for the two tests becomes as shown in Table XVII for Test I, and Table XVIII for Test II.

STEP 2. DETERMINATION OF PERCENTAGE OF PUPILS ABOVE MID-POINTS OF SCORES

The second step is to determine the percentage of pupils above the mid-points of the various scores. As representative of any given score the halfway point may be taken as

TABLE XVII. RANK ORDER OF RAW SCORES OF SIXTEEN PUPILS IN TEST I

Pupil	Score
A	29
B	27
C	26
D	24
E	22
F	22
G	21
H	21
I	20
J	19
K	18
L	17
M	17
N	15
O	14
P	10

TABLE XVIII. RANK ORDER OF RAW SCORES IN TEST II OF SAME SIXTEEN PUPILS AS SHOWN IN TABLE XVII

Pupil	Score
C	57
G	48
I	44
A	42
E	39
B	38
F	35
N	33
K	33
D	30
H	29
M	26
O	23
J	21
L	17
P	10

best indicating the score, since the group may cover a considerable part of a base line. In order to find these percentages the number of individuals who exceed any given score should be added to one half the number of individuals who

have reached that score (in order to locate the halfway point above referred to), and the percentage which that total bears to the entire group of pupils should be determined. The general formula is as follows :

The number of pupils exceeding a given score plus half the number of pupils reaching that score equals what percentage of the total number of pupils?

No. pu. exceeding $+ \frac{1}{2}$ those reaching = $x\%$ of total No.

In Test I, Fig. 25 or Table XVII, no individual in the class exceeded a score of 29. Only one individual received (reached) a score of 29. The total number of pupils in the class was sixteen. Therefore the calculation becomes as follows :

$$0 + \frac{1}{2} = x\% \text{ of } 16.$$

This is found to be 3.13%, which should be placed opposite that score, as shown in Table XIX.

For all ordinary purposes the computations are exact enough when carried out for two decimal places with the second decimal raised if the third decimal would be 0.005 or above, and left unchanged if the third decimal would be below 0.005. In the calculation above, the exact figure to three places was 3.125%, which by the foregoing arbitrary ruling should be raised to 3.13%.

In the same test one individual exceeded a score of 27 and one individual reached a score of 27. Therefore for this score the calculation becomes

$$1 + \frac{1}{2} = x\% \text{ of } 16.$$

This amounts to 9.38%.

Further down on Table XVII it may be seen that six individuals exceeded a score of 21 and two individuals reached it. For this score the computation becomes

$$6 + 1 = x\% \text{ of } 16.$$

This amounts to 43.75%, which can also be tabulated as

in Table XIX, which shows the computations for the percentages of all the scores for Test I, as given in Table XVII.

The other score computations are made in the same way, and the final tabulation of the data for the computations for Test II appears in Table XX on the following page.

TABLE XIX. COMPUTATIONS FOR DETERMINING PERCENTAGE OF PUPILS WHO EXCEED MID-POINTS OF VARYING RAW SCORES OF TEST I

Raw Score	Number of Pupils Reaching Score	Number of Pupils Exceeding Score	One Half of Pupils Reaching Score	Total Amount Exceeding Mid-point	Percentage of Total Number of Pupils
29	1	0	$\frac{1}{2}$	0.5	3.13
27	1	1	$\frac{1}{2}$	1.5	9.38
26	1	2	$\frac{1}{2}$	2.5	15.63
24	1	3	$\frac{1}{2}$	3.5	21.88
22 } 22 }	2	4	1	5.0	31.25
21 } 21 }	2	6	1	7.0	43.75
20	1	8	$\frac{1}{2}$	8.5	53.13
19	1	9	$\frac{1}{2}$	9.5	59.38
18	1	10	$\frac{1}{2}$	10.5	65.63
17 } 17 }	2	11	1	12.0	75.00
15	1	13	$\frac{1}{2}$	13.5	84.38
14	1	14	$\frac{1}{2}$	14.5	90.63
10	1	15	$\frac{1}{2}$	15.5	96.88
Total Number of Pupils, 16					

A quick way of finding percentages. There is a faster method than that ordinarily used for finding percentages when all the percentages have the same base. Expressed simply, this is to find the percentage of one unit of the total and then to multiply by that unit each of the amounts of which the percentage of the total is desired. To do this the following operations are necessary:

1. *Dividing the reciprocal and raising the quotient to percentage.* The reciprocal is the fraction consisting of one divided by the total number for which percentages are to

be found. The reciprocal, for example, in both Test I and Test II is one sixteenth. Dividing this gives, in decimals, 0.0625, which should be multiplied by one hundred (pointing off two places), giving the quotient in terms of percentage, 6.25. This means that one is 6.25 per cent of sixteen.

TABLE XX. COMPUTATIONS FOR DETERMINING PERCENTAGE OF PUPILS WHO EXCEED MID-POINTS OF VARYING RAW SCORES OF TEST II

Raw Score	Number of Pupils Reaching Score	Number of Pupils Exceeding Score	One Half of Pupils Reaching Score	Total Amount Exceeding Mid-point	Percentage of Total Number of Pupils
57	1	0	$\frac{1}{2}$	0.5	3.13
48	1	1	$\frac{1}{2}$	1.5	9.38
44	1	2	$\frac{1}{2}$	2.5	15.63
42	1	3	$\frac{1}{2}$	3.5	21.88
39	1	4	$\frac{1}{2}$	4.5	28.13
38	1	5	$\frac{1}{2}$	5.5	34.38
35	1	6	$\frac{1}{2}$	6.5	40.63
33 } 33 }	2	7	1	8.0	50.00
30	1	9	$\frac{1}{2}$	9.5	59.38
29	1	10	$\frac{1}{2}$	10.5	65.63
26	1	11	$\frac{1}{2}$	11.5	71.88
23	1	12	$\frac{1}{2}$	12.5	78.13
21	1	13	$\frac{1}{2}$	13.5	84.38
17	1	14	$\frac{1}{2}$	14.5	90.63
10	1	15	$\frac{1}{2}$	15.5	96.88
Total Number of Pupils, 16					

2. *Multiplying the totals for each score to find percentage.* The final step is to multiply the numbers in the Total Amount column (see Table XIX or Table XX) by the final figure found in Operation 1, which will give directly a final percentage.

Thus in Table XIX the total amount for score 29 is 0.5. This multiplied by 6.25, the number found in Operation 1, gives 3.125 (or 3.13 per cent) as the percentage desired. Again, in Table XIX the total for a score of 22 is 5.0. This multiplied by 6.25 gives 31.25, the percentage required.

The same procedure is possible throughout all M-scale computations and will save the teacher much time in calculating the percentages.

When using an adding machine the percentages can be found by successive additions of one half the reciprocal, since many amounts are in half steps. Two successive additions will then be necessary to complete a step, three to complete a step and a half, and so on. To make the error as small as possible the reciprocal, after allowance for pointing off, should be carried to five or six decimal places.

STEP 3. FINDING M-SCALE VALUES FOR PERCENTAGES

The next step in M-scaling these tests, that is, in reducing the raw scores on each of the two tests to scores on the same scale, is a simple matter. It is necessary merely to look for the percentage, which is found as a result of the computations in Step 2 (of M-scaling), as given in Table XXI, and to assign to the raw scores in question the scale value opposite the group containing that percentage. The procedure is as follows:

In Test I a raw score of 29 has a percentage of 3.13. By looking in Table XXI it will be found that the scale score for the percentage group which contains 3.13 per cent — the group 2.56–3.21 — is 69. A raw score of 27 on Test I has a percentage of 9.38, and by looking in Table XXI it is found that this lies in the percentage group 8.85–10.55, which gives a scale score of 63, which can then be assigned to a raw score of 27. A score of 18 has a percentage of 65.63, which, by Table XXI, has an M-scale value of 46, which can then be assigned to that score.

Similar procedure can be followed for all the scores on both tests according to the percentages which were found in Step 2; and when this is done, the final M-scale scores for the various raw scores are as found in Table XXII for Test I, and in Table XXIII for Test II.

TABLE XXI. M-SCALE VALUES FOR PERCENTAGES FOUND IN STEP 2

PERCENTAGES		PERCENTAGES	
Between	Value	Between	Value
0.0026-0.0038	90	48.010-51.98	50
0.0039-0.0058	89	51.990-55.95	49
0.0059-0.0089	88	55.960-59.86	48
0.009-0.0129	87	59.870-63.67	47
0.013-0.018	86	63.680-67.35	46
0.019-0.027	85	67.360-70.87	45
0.028-0.039	84	70.880-74.21	44
0.040-0.057	83	74.220-77.33	43
0.058-0.081	82	77.340-80.22	42
0.082-0.10	81	80.230-82.88	41
0.110-0.15	80	82.890-85.30	40
0.160-0.21	79	85.310-87.48	39
0.220-0.29	78	87.490-89.43	38
0.300-0.39	77	89.440-91.14	37
0.400-0.53	76	91.150-92.64	36
0.540-0.70	75	92.650-93.93	35
0.710-0.93	74	93.940-95.04	34
0.940-1.21	73	95.050-95.98	33
1.220-1.57	72	95.990-96.77	32
1.580-2.01	71	96.780-97.43	31
2.020-2.55	70	97.440-97.97	30
2.560-3.21	69	97.980-98.41	29
3.220-4.00	68	98.420-98.77	28
4.010-4.94	67	98.780-99.05	27
4.950-6.05	66	99.060-99.28	26
6.060-7.34	65	99.290-99.45	25
7.350-8.84	64	99.460-99.59	24
8.850-10.55	63	99.600-99.69	23
10.560-12.50	62	99.700-99.77	22
12.510-14.68	61	99.780-99.83	21
14.690-17.10	60	99.840-99.885	20
17.110-19.76	59	99.886-99.917	19
19.770-22.65	58	99.918-99.941	18
22.660-25.77	57	99.942-99.959	17
25.780-29.11	56	99.960-99.971	16
29.120-32.63	55	99.972-99.980	15
32.640-36.31	54	99.981-99.986	14
36.320-40.12	53	99.987-99.990	13
40.130-44.03	52	99.991-99.994	12
44.040-48.00	51	99.995-99.996	11

Table XXI is an adaptation of the table used by Dr. McCall in his T-scale construction, which is given in its original form in his *How to Measure in Education*.¹ The

TABLE XXII. M-SCALE VALUES
FOR RAW SCORES ON TEST I

Raw Score	Per Cent	M-Scale Value
29	3.13	69
27	9.38	63
26	15.63	60
24	21.88	58
22	31.25	55
21	43.75	52
20	53.13	49
19	59.38	48
18	65.63	46
17	75.00	43
15	84.38	40
14	90.63	37
10	96.88	31

TABLE XXIII. M-SCALE VALUES
FOR RAW SCORES ON TEST II

Raw Score	Per Cent	M-Scale Value
57	3.13	69
48	9.38	63
44	15.63	60
42	21.88	58
39	28.13	56
38	34.38	54
35	40.63	52
33	50.00	50
30	59.38	48
29	65.63	46
26	71.88	44
23	78.13	42
21	84.38	40
17	90.63	37
10	96.88	31

construction of the original table is there described in full,² and the caption on the table reads:

TABLE 23

Shows the S.D. distance of a given per cent above zero. Each S.D. value is multiplied by 10 to eliminate decimals. The zero point is 5 S.D. below the mean.³

Table 23, referred to, and Table XXI, here, are based upon plus and minus 5 S.D.⁴ of a normal curve. The object

¹ Chaps. ix and x.

² Ibid. p. 273.

³ Ibid. pp. 274-275.

⁴ Ibid. pp. 383-386. "S.D." is the abbreviation for Standard Deviation, or, as the unit is sometimes called, the Mean Square Deviation. It is much used in statistics as a measure of variability.

is merely to find some standard to which the scores on tests can all be reduced so that the tests may be made comparable in their units. The "Values" given in Table XXI are merely units on a base line in terms of one-tenth S. D. distance with the zero point at minus 5 S. D. to eliminate negative quantities. Fig. 27 gives the relations of the two scales. From this it can be seen that an M-scale point of 30 is equivalent to minus 2 S. D., or of 83 to plus 3.3 S. D., and the like.

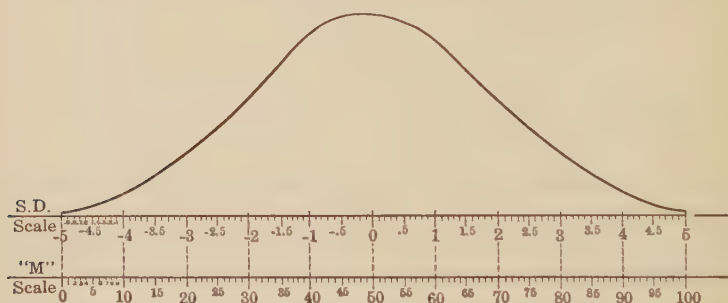


FIG. 27. Relation of S. D. Scale and M Scale

The table as used by Dr. McCall gives S. D. values from zero to one hundred in intervals of five tenths. It is adapted for use with large numbers of cases, and for purposes of M-scaling it can be simplified as in Table XXI, which gives the S. D. values in single units of the same size as those of Dr. McCall, from eleven to ninety, which is a larger range even than is likely to be used for these Teacher's Tests if the directions given in these chapters have been followed. If a greater range is needed for any reason, the original table as given by Dr. McCall should be consulted.

Making composite of M-scale scores on two or more tests. So long as a teacher has the original papers in front of him during the foregoing calculations, and is thereby enabled to refer a raw score back to the papers for identification of the pupil making that score, it is unnecessary to make any permanent record of the raw scores, since they are no longer

needed. When each set of M-scale scores has been found, however, it should be carefully copied in the roll book or other permanent record of the teacher, because he has later to depend upon these scores in rating, promotion, or classification. The calculations shown in Tables XIX and XX, and

TABLE XXIV

Pupil	Test I Raw Scores	M-Scale Value	Test II Raw Scores	M-Scale Value
A	29	69	42	58
B	27	63	38	54
C	26	60	57	69
D	24	58	30	48
E	22	55	39	56
F	22	55	35	52
G	21	52	48	63
H	21	52	29	46
I	20	49	44	60
J	19	48	21	40
K	18	46	33	50
L	17	43	17	37
M	17	43	26	44
N	15	40	23	42
O	14	37	23	42
P	10	31	10	31

Scores on Tests I and II are derived from Tables XVII and XVIII. M-scale scores are derived from Tables XIX and XX.

in Tables XXII and XXIII, need not be preserved either, as they are merely steps in the computation of the M-scale scores. The form of tabulation of the M-scale scores is immaterial provided the teacher keeps the record so that each pupil's scores may be easily found when wanted.

In making a composite the teacher should add together all the M-scale scores for each individual pupil on all the tests which are to be used in the composite. This is an easy matter if the only score records are the M-scale scores,

and there is no danger, therefore, of confusion with raw scores or other temporary calculations.

Table XXIV, which shows the first step in making the composite, can be eliminated by the teacher if he simply transfers his records from the individual papers to the roll

TABLE XXV. TABULATION OF COMPOSITE SCORES FROM M-SCALE SCORES

Pupil	M-Scale Score Test I	M-Scale Score Test II	Sum of Scores	Composite Sum divided by Number of Tests
A	69	58	127	64
B	63	54	117	59
C	60	69	129	65
D	58	48	106	53
E	55	56	111	56
F	55	52	107	54
G	52	63	115	58
H	52	46	98	49
I	49	60	109	55
J	48	40	88	44
K	46	50	96	48
L	43	37	80	40
M	43	44	87	44
N	40	50	90	45
O	37	42	79	40
P	31	31	62	31

This might be considered a sample page from the roll book of the teacher, containing only pupils' names, M-scale scores, and the final composites.

book. It is a necessary step in this explanation, however, because it shows the scores made by the pupils on the separate tests.

Table XXV shows the next step in making the composite and may be considered a page from the roll book of the teacher. The first column gives the names of the pupils, which are here indicated by the same letters which have been used previously; the two following columns give the

M-scale scores as found in the preceding steps; the fourth column gives the addition of the M-scale scores in preparation for the composite; and the last column gives the final composite score for each pupil, which in this case was found by taking the sum of the two M scores, as found in the fourth column and dividing it by two, the number of tests entering the composite.

If there are three or more tests, the procedure is just the same: The scores for all the tests are M-scaled, and in order to find a composite of all the M-scale scores they are added together and divided by the number of tests which have entered into each composite. The result is a composite in which all the elements entering into the final scores have equal weighting. The formula for finding a composite, regardless of the number of tests entering into the composite, is as follows:

$$\frac{\text{Sum of M scores}}{\text{Number of tests}} = \text{Composite}$$

The next chapters are devoted to the use of these composite scores for rating, promoting, and classifying pupils, but it would be well for the teacher to keep in mind the fact that an M score of 50 indicates the average of the class in a single test, that likewise an M-scale composite of 50 indicates the average of the class in the sum of the tests given, that M-scale and composite scores above 50 indicate scores above the average, and that M-scale scores below 50 indicate the scores below the average.

Chapter summary. A reliable composite score for pupils on several tests is frequently a desirable goal for a teacher, but the mere addition of the raw scores on the several tests will not give a fair composite. This can be found only by reducing the raw scores to scores on the same scale, and a method is here described, called M-scaling, which reduces scores to a standard unit for finding a composite. Scores found in this way have a constant interpretation.

The method involves finding the percentage of pupils who attain each score on the individual tests and, by means of a uniform table based on uniform distances on the base line of the frequency distribution of a normal group (Table XXI), transmuting these percentages into standardized M-scale scores. These M-scale scores, being of equal value throughout, are then added together and divided by the number of tests which enter them, to form the desired composite scores for the class group in question.

SELECTED BIBLIOGRAPHY

- MCCALL, W. A. *How to Measure in Education*, chaps. ix and x. The Macmillan Company. New York, 1923.

CHAPTER XV

SPECIFIC USES: JUDGING PUPILS IN ACHIEVEMENT ACCORDING TO ABILITY

The use of tests for rating. As far as rating or marking is concerned, the relation of tests to grades has already been discussed in Chapter X. There it was stated that grades should be given in one of two ways: either in relation to the achievement of the pupil in terms of his ability to achieve or else in relation to the achievements of the group of which he is a part. This latter seems to be the prevailing scheme of determining grades at the present time, and the grade given, therefore, is not so much a measure of a pupil's own efforts as it is of his place in the class group. Both types of ratings are of value, however, the one in the placement and promotion of pupils and the other in the judgment of the quality of work or effort of pupils, and, if possible, both types of grades, or ratings, should be used. It should be easy to see, for example, that if pupils were marked upon the degree of their achievements in relation to their capacities to achieve, a pupil with a relatively low capacity and a somewhat relatively higher achievement would deserve a somewhat higher mark, or grade, *in effort* than a pupil with a high ability and a relatively less achievement. In this case the total achievement of the less capable pupil might be less than the total achievement of the more capable pupil, but his rating nevertheless should be higher in effort. A technic for determining this type of effort or accomplishment rating, called the Accomplishment Ratio Technic, has been devised for use with Standard Tests, and it is therefore possible for the interested teacher to find a grade, or mark, based upon the ability of a pupil to achieve in Standard Tests.

The Accomplishment Ratio Technic in Standard testing.¹ The first step in the development of the Accomplishment Ratio Technic in Standard testing consists in the determination of the absolute abilities of pupils. This is found by giving some Standard group or individual intelligence test and from that deriving the Intelligence Quotients of the several pupils. From this Intelligence Quotient (usually known as I. Q.), in connection with the known Chronological Age of the pupil, the Mental Age of the pupil can be found by the formula $IQ \times CA = MA$. This Mental Age is the age at which pupils should be, from an intellectual standpoint, if they have made full use of their inherent abilities.

The second step in the process when used with Standard Tests is the derivation of the Educational, or Subject, Ages of the pupils. These are comparable to the Mental Ages, but are obtained from Standard educational tests. Since these tests have the *norms* of the answers, that is, the actual results attained by other pupils all over the country on these same tests, or else by similar types of pupils on the same tests, and since the average age of pupils throughout the country is well known through widespread investigation, it is possible by means of standardized modes of transposition to turn the test scores into Educational Ages that are known as Arithmetic Age, Spelling Age, and the like. These Educational Ages are merely the ages at which other pupils in general accomplish a like amount of the test materials. They are, in effect, representative of the actual present educational status of a pupil.

The third step in the process is to compare directly the two ages thus found: first, the age at which a pupil should be if he had done all that he might; secondly, the age at which he is actually found to be. This comparison is direct between

¹ Raymond Franzen, *The Accomplishment Ratio* (Teachers College Contributions to Education No. 125) (Teachers College, New York City, 1922); W. S. Monroe and B. R. Buckingham, *Teacher's Handbook*, Illinois Examination (Public School Publishing Company, Bloomington, Illinois, 1920).

the two ages, in terms of a ratio. Since the Mental Age is the determining factor of the ratio, it becomes the denominator of the fraction which represents the ratio; and since the ratio is to be a measure of the success of the pupil in reaching his intellectual possibilities, the Educational Age becomes the numerator. Thus the formula for A. R. (Achievement Ratio) becomes

$$\frac{\text{Educational Age}}{\text{Mental Age}} = \text{Achievement Ratio},$$

or as usually written symbolically,

$$\frac{EA}{MA} = AR.$$

In practice the Achievement Ratio is carried out to two decimal places and raised by 100 to eliminate the decimal point. This is not necessary, however, and some writers and teachers prefer to use the ratio as originally found. A ratio of 1.00, however, is the same as one of 100 given by another writer, the only difference being the use of the decimal point. When these ratios are to be used by pupils below the school grades where decimals are taught, it is perhaps wise to use the latter type of rating. Thus if a pupil has done all that he should, his E.A. and M.A. are the same and his A.R. is 100. A pupil who has a Mental Age of one hundred and forty-four months has an Educational Age of one hundred and forty-four months and by the formula his Achievement Ratio is 100:

$$\frac{144(EA)}{144(MA)} = 100(AR).$$

An Achievement Ratio of 100, then, means that a pupil, regardless of his respective place in a class group, is accomplishing all that may reasonably be expected of him.

If, however, a pupil has an Educational Age of, for example, one hundred and fifty-five months, which is greater than his Mental Age, say one hundred and forty-four months, it

means that he has done better than there has been reason to expect, since he has exceeded the achievement to which his ability should normally bring him. This is shown by an A. R. of more than 100 :

$$\frac{155(EA)}{144(MA)} = 108(AR).$$

Such a pupil should be praised for what he has done, but should not be forced to put forth greater efforts, regardless of where he is placed with respect to the class group, because he is already exceeding normal expectations for a pupil of his ability.

Again, if a pupil with an Educational Age of, for example, one hundred and thirty months has a Mental Age which is greater, say one hundred and forty-four months, it means that he has done less well than should be expected of him, because he has not maintained an intellectual level equivalent to his possibilities as shown by his A. R., as follows :

$$\frac{130(EA)}{144(MA)} = 90(AR).$$

An A. R., then, of less than 100 means that such a pupil should be prodded, or encouraged, or diagnosed to find out the what and the why of his difficulties.

The essentials of the ratio technic. In order to apply this technic for use in Classroom Tests some other scale than an age scale is necessary, and therefore it becomes necessary to inquire into the essentials of this technic to discover, since the specific elements cannot be used, the essentials that must be met. These specific elements, such as Educational Age, are not available from Classroom Tests, because Classroom Tests have no *norms* (are not standardized), and these ages are derived from these norms.

The first essential of the A. R. Technic is a fixed and standard rating of mental ability. In the Standard Test Achievement Ratio Technic this is derived, as has been

shown, from Standard Intelligence Tests and later changed into a Mental Age that gives a measure of where the pupil ought to be.

The second essential in this technic is a group of test scores or results, which are derived from Standard Tests and are then converted into Educational Ages. This gives a measure of present educational status.

The third essential, and this is of paramount importance, is that these two ratings, mental and educational, shall be in exactly the same kind of unit, so that they may be directly comparable with each other. When they are so comparable, placing them in a ratio will bring a result which, if treated as above, will give ratings above, below, or at 100 with the interpretations previously outlined.

The establishment of a ratio technic for use with Teacher's Classroom Tests. If an Accomplishment Ratio Technic were possible for use with these Classroom Tests, as it is with Standard Tests, the usefulness of the results of such tests would be considerably amplified. It would mean that, as a result of such tests, pupils could be told the extent of their achievements in relation to their abilities, and the teacher could know not only exactly which pupils were in need of stimulation but also those pupils whose work was satisfactory.

Since the two ratings needed for such a technic are, first, a rating of ability and, secondly, a rating of achievement, it follows that two such ratings must be found for the results of Classroom Tests. Since, also, these two ratings must be comparable, whatever two ratings are found must be in terms of the same kind of unit.

The one uniform rating that has already been found consists of the M scores and M composite scores. These are satisfactory to use as numerators in fractions for Achievement Ratios, and therefore all that is needed is an M-scaled rating of ability. Such a rating of ability is possible through the use of good group intelligence tests. If a good group intelligence test be given to the pupils who are

represented in the Classroom Tests, the intelligence-test raw scores (the total point scores received by each pupil), or any other rating of these tests which shows the relative relationship of the members of the group taking the test, can be used in M-scaling exactly as in the case of the Classroom Tests. These M-scaled intelligence scores thus become standard M-scale scores which are comparable both in derivation and in units to the M scores derived from the Classroom Tests. If, then, these scores, with the reservations noted later, be considered as standard measures of ability and the Classroom Test M scores be considered measures of achievement, a direct Achievement Ratio can be found by comparing the two scores exactly as in the case of Standard Tests. The formula would be as follows:

$$\frac{\text{M score in achievement}}{\text{M score in ability}} = \text{Achievement Ratio,}$$

$$\frac{EM}{AM} = AR.$$

1. FINDING ACHIEVEMENT RATIOS FROM THE RESULTS OF A SINGLE CLASSROOM TEST

STEP 1. FIND M-SCALE TEST SCORES

In order to find an Achievement Ratio from the results of a single Classroom Test (or battery of tests) the first step is to derive the varying M scores for that test, as was described on pages 270-278 of Chapter XIV.

STEP 2. DERIVE THE STANDARD M SCORES

The second step is the derivation of what will be called from now on standard M-scale scores, or the scores that can be used as ability ratings.

As was suggested, these should be derived from Standard group intelligence tests. Since these are Standard Tests,

from which Intelligence Quotients can be obtained, and since we know that Intelligence Quotients carefully derived from reliable tests are relatively constant for the individuals concerned,¹ it can be assumed that if the M scores derived from these tests are carefully found, they will be justly representative of the range and relative relationships of the varying abilities within a class group, *so long and only so long as it remains a constant group*. It would be especially true that these M scores of ability, other things being equal, would be constant during the single year that most teachers have a single group of pupils. Thus for any group of pupils it should be sufficient for a teacher to determine this standard M score just once during the school year, checking it of course by any means available, such as the M-scale ratings of other Standard Tests that may be used. It should be possible, then, after having once determined the standard M ratings for a class, to assign to the various pupils their standard M scores for that semester or year, to be used in determining their degrees of accomplishment during that time.

In finding this standard M score the following is the procedure: The teacher should first of all select and give a Standard Group Test of Intelligence.² An individual test, such as the Stanford revision of the Binet-Simon Intelligence Scale, can be used for some pupils, if need be; but if two tests are thus used, one for some pupils and another for others, it will be necessary for the teacher to reduce both sets of scores to a single type of score, usually a Mental Age, before following the procedure given here.

Assuming that the teacher has selected a single group test to be used, he may M-scale his results according to the type of score he secures from the tests. If the results

¹ Compare M. R. Trabue's *Measuring Results in Education*, pp. 416, 435. American Book Company, New York, 1924.

² See W. A. McCall's *How to Measure in Education*, pp. 78-79, and M. R. Trabue's *Measuring Results in Education*, pp. 425-426, for advice on the selection of a suitable Intelligence Test.

from the tests are in terms of point scores, as in the National Intelligence Test, these raw point scores can be M-scaled directly, as outlined in the preceding chapter. The process is the same if the results are in terms of Mental Ages. The teacher should remember, however, that *Intelligence Quotients cannot be M-scaled*. If a teacher is so fortunate as to have a complete set of reliable Intelligence Quotients for his class, these quotients must first be turned into Mental Ages before being M-scaled.

The raw point scores or the Mental Ages, as the case may be, should first be placed in rank order of from greatest to least and treated *exactly* as if they were raw scores on a Classroom Test. The number of pupils exceeding each score should be determined and added to one-half the number of pupils who receive or reach that score, and the percentage of that total to the entire number of pupils should then be determined. With this percentage determined, the final step is to find from Table XXI, p. 276, the M-scale value of each percentage found, which becomes the standard M score for the pupil reaching the score from which that value was determined.

The teacher should have no particular difficulty in giving and scoring the intelligence test which is used, and there is no good reason why the M-scale ratings that are received should not be valuable. A few precautions should, however, be observed by all teachers.

1. *Precautions in selecting a test.* The teacher should select a test adapted to his pupils. This may be any test which the teacher may know or may desire to use, if it complies with this condition. For the inexperienced examiner the reference in footnote 2, p. 289, from Dr. McCall should be of distinct help.

2. *Precautions in giving individual tests.* Unless the teacher is a trained expert, no individual test of intelligence should be used by him. If for any reason an individual test seems necessary as a check on some pupil, it should be given by

someone who is familiar with the tests and is an expert in giving them. If such tests are given, or if Intelligence Quotients are available for only a part of a class, the scores for all the class should be converted into Mental Ages before M-scaling is attempted.

3. *Precautions in giving the group tests.* The tests that are used should be given *exactly* according to the directions which accompany them. The teacher should be extremely careful to observe each of these directions faithfully and should practice giving the tests in private before attempting to give them to his pupils. Unless the directions which accompany the test specifically permit it, the teacher should give his pupils no help whatsoever in doing the test *beyond the actual words* of the directions. In these tests the teacher must remember that the difficulty of the test is a part of the test.

4. *Precautions in scoring the test.* In scoring the test the teacher should observe *exactly* the directions which accompany the test. If there is any doubt at all as to what is the proper score on such a test (such doubt may occur in exceptional cases only), such as whether a dollar sign is necessary for a correct answer or whether an answer can be written either as a fraction or as a decimal, the teacher should exercise his best judgment as to what was meant and be consistent according to that judgment throughout the entire scoring of that particular element.

No pupil should, after a test is finished, be told what was a correct answer for any particular element, nor should any pupil be given his point rating or his Intelligence Quotient. These should be kept *absolutely* private and should not be accessible to any but properly constituted school authorities. It is professionally unethical to do otherwise. In giving the standard M scores to pupils they can be told that that is the rating which they are expected to get in future papers.

5. *Precautions for special re-tests.* It may be that for some reason a teacher is doubtful as to the fairness of the test

for some particular pupil or pupils. There are certain conditions which make it impossible for an individual to do his best, that is, show his true ability, on these tests. Headaches or illness of any sort, or excessive fatigue at the time of taking the test (or excessive excitement, which amounts to the same thing in the end), or the like may be conditions of this character. In such a case the teacher should give another test, preferably a different form of the same test that was given first. If this is done the highest score obtained on *either* test should be taken as basic, for the reason that it should be supposed that an individual cannot at any time do better than his best. If there are scores for the class group from two tests of this sort, unless the raw scores on both tests are equivalent point for point the teacher should convert all the scores for all pupils into Mental Ages before M-scaling. The procedure for this is as follows:

a. Convert point scores into Intelligence Quotients.

b. Multiply Intelligence Quotients by Chronological Ages of pupils. This gives a Mental Age.

6. *Precautions to be observed when pupils enter or leave a class group during a semester or year.* When pupils enter or leave a class, the class group changes slightly and the standard M scores will also change somewhat. Any change is of course important, but unless the shifting of the members of the class is excessive, it is unlikely that a few changes will change the character of the group greatly.

Each pupil who enters the class, after the first group intelligence test has been given, should be given the *same intelligence test* as was given to the other members of the class and, for temporary purposes only, should be assigned a standard M score which was previously assigned to the score nearest to his own. This will probably work very little injustice in Achievement Ratios.

When a pupil leaves a class the teacher, temporarily at least, should make no changes in the scores.

If the character of the class group changes markedly, the teacher can always find new standard M scores for the pupils by M-scaling the intelligence-test scores of the pupils again, adding to the original group of scores those of the new pupils who have entered and eliminating from that group the scores of the pupils who have left. If this changes the standard M-scale ratings of some pupils from their earlier scores, the teacher should remember that M-scaling gives the rating of an individual with respect to the group in which he is (it is not an absolute rating like that of the Mental Age) and that, therefore, it is not the pupil but rather the group that has changed.

2. FINDING ACHIEVEMENT RATIOS OF PUPILS WHEN A COMPOSITE OF SEVERAL TESTS IS USED

The suggestions given in the foregoing paragraphs relate to the process of finding Achievement Ratios of pupils with respect to a single test. When a teacher wishes to find the Achievement Ratio of a pupil in the composite of a number of tests, there is a slight addition to the previous technic which is necessary. The standard M-scale scores of ability are exactly the same as used with the single tests, but the composite, as found in Chapter XIV, is not exactly comparable to it. The range of the Composite is shortened (or attenuated) from that of any of the single tests used.

It will be noticed, for example, that in Table XXV of Chapter XIV (p. 280) the M-scale scores of Pupil C are 60 and 69 respectively on Tests I and II and that Pupil C has a composite rating of 65. It should also be noted that this composite rating is the highest rating for the class in the composite. If it be supposed that Pupil C has a standard M score of 69, as he would have if he had the highest ability score in a group of sixteen pupils, then for Test I he would have an A.R. of 87, and on Test II his A.R. would be

100, both of these A.R.'s being determined as above. On his composite score, however, Pupil C retains his leadership in the class group and so has been satisfactory in terms of an Achievement Ratio of this character. Nothing but an M-scale composite rating of 69 will show this (an A.R. of 100),

TABLE XXVI. REVISED M-SCALING OF COMPOSITE SCORES GIVEN IN
TABLE XV, CHAPTER XIV

Pupils	M-Scale Composite	Number of Pupils Reaching Score	Number of Pupils Exceeding Score	One-Half Number of Pupils Reaching	Total Amount Exceeding Mid-point	Percentage of Total Number of Pupils	Final Revised M-Scale Com- posites
C	65	1	0	$\frac{1}{2}$	0.5	3.13	69
A	64	1	1	$\frac{1}{2}$	1.5	9.38	63
B	59	1	2	$\frac{1}{2}$	2.5	15.63	60
G	58	1	3	$\frac{1}{2}$	3.5	21.88	58
E	56	1	4	$\frac{1}{2}$	4.5	28.13	56
I	55	1	5	$\frac{1}{2}$	5.5	34.38	54
F	54	1	6	$\frac{1}{2}$	6.5	40.63	52
D	53	1	7	$\frac{1}{2}$	7.5	46.88	51
H	49	1	8	$\frac{1}{2}$	8.5	53.13	49
K	48	1	9	$\frac{1}{2}$	9.5	59.38	48
N	45	1	10	$\frac{1}{2}$	10.5	65.63	46
J	44	2	11	1	12.0	75.00	43
M	44						
L	40						
O	40	2	13	1	14.0	87.50	38
P	31	1	15	$\frac{1}{2}$	15.5	96.88	31
Total Number of Pupils, 16							

and it would be unfair, both in his case and in the case of other pupils, to make his A.R. less than 100, as would happen if the A.R. were derived from an achievement composite of 65 and a standard M of 69.

In order to correct this, when composite scores are to be used for Achievement Ratios the composite scores of all the tests should be M-scaled just as though they were original point scores. This will bring the range of the composite ratings back to a range comparable to that of a single test.

It will correct the attenuation in the composite M ratings caused by the averaging of individual M-scale ratings of the members of a group whose relative placings to one another have changed slightly from one test to another. Table XXVI shows the revised M-scaling of the composite given in Table XXV, Chapter XIV.

Interpretations and values of Achievement Ratios. Achievement Ratios obtained as has been outlined are subject to interpretation similar to those mentioned in the discussion of this ratio in connection with Standard Tests. These are substantially as follows:

1. An Achievement Ratio which is greater than 100 indicates a pupil who is doing better than should be expected of one of his ability, regardless of whether his relative placement in his class group is low or high.

Thus, in the case just cited, if Pupil N had a standard M rating of 38, his A.R. would be

$$\frac{46(EdM)}{38(AbM)} = 121(AR).$$

In spite of the fact that Pupil N is below the average of the class (bearing in mind that an M rating of 50 is average), he is doing considerably better work than we have reason to expect. Such a pupil should not be urged to do more and should be congratulated on what he has done. This is a good illustration of what investigators are actually finding as a result of using the Accomplishment Ratio Technic, — that the pupils who are less capable have been prodded and pushed and pulled to an achievement above their normal level (an achievement that should be satisfactory), whereas pupils of higher ability have been consistently allowed to slump, because the point to which they have slumped is yet above that of their fellows.¹

2. An A.R. of 100 or thereabouts may be considered a satisfactory achievement in terms of ability. If Pupil E

¹ Compare M. R. Trabue, op. cit. p. 446.

has a standard M score of 55, his A.R. may be considered satisfactory. It is found as follows:

$$\frac{56}{55} = 102.$$

Pupil E is doing good work and he should be encouraged to continue in the same way.

3. An A.R. of less than 100 indicates that a pupil is doing less than we should expect of one of his ability, and his case should be investigated for diagnosis and immediate remedial attention. If Pupil I, with a standard rating of 63, gets an educational composite M rating of 54, his A.R. is considerably below 100 and he is doing much more poorly than we should expect from one of his ability, even though he is doing better than the average of the class. For example:

$$\frac{54}{63} = 86.$$

The teacher should immediately consider this case and try to find the reasons for the low Achievement Ratio. A clue to Pupil I's difficulties may be found in some one of his tests, indicating a particular difficulty in some special phase of his work. As will be seen by referring to Table XXV, Chapter XIV (p. 280), Pupil I received an M score of 49 on Test I and of 60 on Test II. It is evident that his difficulty lies far more in the subject of Test I than in that of Test II; therefore it is to that subject that the teacher should direct attention to find the pupil's difficulties. In other cases it may be found that the pupil is not interested or even is lazy. Laziness, brought about by the power to "get by" because of high relative ability, can very easily be detected and brought to light by the use of Achievement Ratios.

What has already been said should have emphasized the values of Achievement Ratios. One of the chief values of the ratio is that it enables a teacher to keep a constant check on the achievements of his several pupils in terms of their

abilities. Although the Achievement Ratio obtained from Standard Tests is more valuable in many ways because of the more significant data upon which it is based (since from it a teacher cannot only determine the degree of acceptability of past achievement of a pupil but also form a basis for estimating what his future achievement ought to be¹), yet the Achievement Ratio obtained from Classroom Tests has its place in enabling a teacher to keep a more constant check on his pupils. This is because teachers are likely to give Classroom Tests more often than they are to give Standard Tests and because they are likely to give Classroom Tests in a wider range of subjects than they do Standard Tests.

It is for this reason that a teacher should find the Achievement Ratios of the pupils on every test that is given, and it is also the reason why, in the end, the Achievement Ratios on the separate tests will be more valuable for everything except promotion ratings than will the Achievement Ratios of the composites of the test scores.

Another value of the Achievement Ratio is that it enables a pupil to keep a constant check on his own work; in this way the ratio aids in the motivation of school work, by providing desirable objectives, as well as in the maintenance of interest. For this reason each pupil should be encouraged to keep a record of his Achievement Ratios, and each pupil, too, especially in and above the fifth grade, should be taught to find them for himself. It is a simple matter to teach these pupils how to determine this rating, and it will be to the decided advantage of the teacher, in point of economy of time and labor, to do so. A simple record card, which each pupil can keep for himself and add to as each test is taken, is suggested below. This record card can be kept by the pupil and checked by the teacher from time to time. Pupils in the Detroit public schools keep similar records of their achievement in various school subjects and seem greatly to enjoy doing so.

¹ Compare W. A. McCall, *op. cit.* chap. iii.

Such a card can be kept for each subject or for all subjects together, as the teacher may choose. The criterion should be to keep the line of Achievement Ratios as nearly on the 100 line as possible, or above it.

Not the least valuable of the advantages of this Ratio Technic is the insight which it gives a teacher into the workings of a class. To know that a pupil is doing good work is

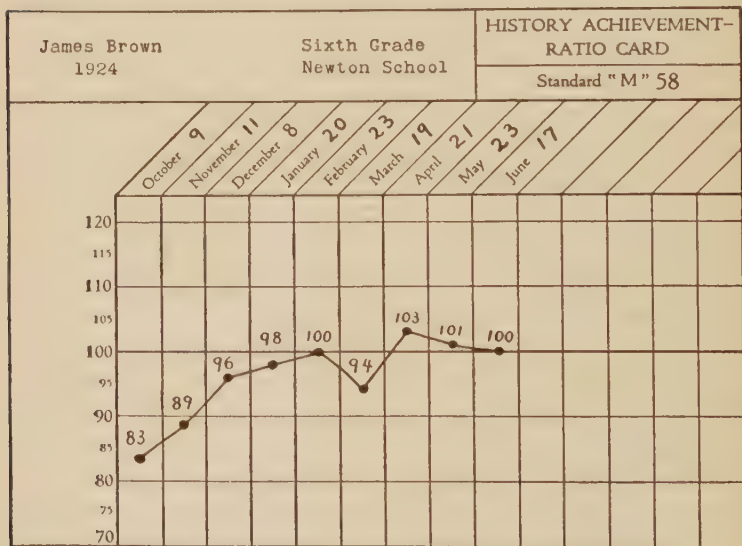


FIG. 28. Record Card for Achievement Ratios

something, but to know in addition that, though his work is good, it is not good enough, is considerably more. To be able to tell a pupil that his work is not the best he can do, and, what is more, to have him *prove* it to himself, furnishes a new leverage in teaching which should not be ignored. To praise the work of a pupil who is doing work below the average of the class will be a new and a delightful experience to many teachers, and receiving this praise will be a similar delight to every such pupil. Criticism of the work of a good pupil, in

spite of its goodness, will come as a surprise to such a pupil, and if it does nothing more it will at least start thought.

Chapter summary. A way of measuring achievement in relation to ability in Standard testing is called the Accomplishment Ratio Technic. It consists in comparing a test result from a Standard educational test with a similarly derived unit which is obtained through giving an intelligence test, and which, therefore, is a measure of ability.

The technic can be adapted to the use of results from Classroom Tests by M-scaling the point scores or the Mental Ages of a class group in a group intelligence test and comparing the results, as standard M-scale scores, with M scores on the educational Classroom Tests.

When the M scores are formed into a composite, the technic must be extended one step in order to make the composite scores comparable with the M-scaled intelligence-test scores. This extra step is merely that of M-scaling the M-score composites of several tests, so as to bring the final M scores within the same range as the standard M scores.

The comparison to find a ratio is always a fraction consisting of the standard M-rating as a denominator and the M-rating of an educational test, here a Classroom Test or composite of Classroom Tests, as a numerator. The final ratio is a figure, with decimal points removed after being carried two places, which is either above, at, or below 100.

The interpretation of Accomplishment Ratios is simple. Ratios in excess of 100 indicate efforts in excess of normal; ratios of 100 indicate normal and expected achievement in terms of ability; ratios below 100 indicate achievement less than could be expected of that ability.

Briefly the values of the Achievement Ratio are that it enables a teacher to give a rating to pupils in terms of their abilities rather than merely in terms of their total achievements; that it enables a teacher to keep a constant check on pupils in terms of what those pupils might be expected to do; that its use helps to provide a worthy motivation for school

work and thereby promotes interest and attention; and finally, that it gives a teacher a new insight into the development of his classes.

SELECTED BIBLIOGRAPHY

- FRANZEN, R. The Accomplishment Ratio (Teachers College Contributions to Education No. 125). New York, 1922.
- MCCALL, W. A. How to Measure in Education, pp. 85-87, 154-155, 210-226. The Macmillan Company, New York, 1923.
- PINTNER, R. Intelligence Testing (chap. vi, "Group Tests"). Henry Holt and Company, New York, 1923.
- GREGORY, C. A. Fundamentals of Educational Measurement (pp. 122-136, "Group Intelligence Tests"). D. Appleton and Company, New York, 1922.
- TRABUE, M. R. Measuring Results in Education, pp. 419-427. American Book Company, New York, 1924.
- BROOKS, S. S. Improving Schools by Standardized Tests (chap. ix, "Some Uses for Intelligence Tests"). Houghton Mifflin Company, Boston, 1922.
- STEVENSON, P. R. Smaller Classes or Larger. Public School Publishing Company, Bloomington, Illinois, 1924.
- BUCKINGHAM, B. R., and MONROE, W. S. "A Testing Program for Elementary Schools," in *Journal of Educational Research*, Vol. II, pp. 521-532.
- MONROE, W. S. The Illinois Examination. University of Illinois Bulletin, *Bureau of Educational Research Bulletin No. 6*, Vol. XIX, No. 91 (October 31, 1921).

CHAPTER XVI

SPECIFIC USES: JUDGING AND CLASSIFYING PUPILS ACCORDING TO GROUP PLACEMENT

The relation of tests to rating, promotion, and classification. Whether a pupil should or should not be promoted to the next upper class, how he should be marked in a school system where the success of a pupil is measured in terms of certain grades or marks, and how to achieve a classification in a school have been in great measure, in the past, matters of judgment by the classroom teacher. The judgment of the teacher has been in most cases good, in some cases even excellent; and in all cases the degree of excellence has been directly in accordance with the degree to which the judgment of the teacher and the actual facts have coincided. There are many factors having to do with all these phases of school work, and teachers have taken them into consideration. One has been the judgment of the teacher with respect to the kind of school work which the pupils have done or the degree of school achievement which they have attained. Here the judgment of the teacher has been reënforced by the examinations or tests which the pupils have been given, but other factors have entered as well. The maturity of the pupil, or rather the teacher's judgment of his relative maturity, has been one factor which has been considered. Another factor has been the general attitude of the pupil. Has he been quick and eager, or has he been indifferent and slow? These are all legitimate factors in the marking, promotion, or classification of pupils, and they have been used with success.

It is undoubtedly true, on the other hand, that other factors have entered into this judgment and have perhaps carried undue weight. Is the child clean and orderly, or has

he been slovenly? We do not yet know that there is any positive relation between slovenliness and inability to do school work, though we may suspect that the child who is slovenly in dress and habit is also slovenly in thought. Even here, however, it is possible for most of us to find outstanding examples of men or women who belie this seeming correlation. Is the child who speaks with hesitation necessarily not sure of his facts? In many of our own groups of friends it is possible to find the glib and careless speaker and the brilliant person who is hesitating and almost silent. Does a bright and shining face unfailingly indicate intense interest and capacity for high scholarship? Doubtless there is usually a close connection, since it has been accepted as an indication by teachers from time immemorial; but the brightest individual, from the standpoint of intelligence, that the writer ever encountered seemed always on the point of going to sleep, whereas a bright-eyed, clean, and wholesome-appearing boy of his acquaintance once failed miserably to establish the right to even normal intelligence.

Testing, and especially the type of testing that has been described in the foregoing chapters, is a good substitute for otherwise unsubstantiated opinion. We, as teachers, need all the objective evidence possible in our decisions with respect to our pupils, and every effort should be made to support objective evidence with our better judgment rather than to support our judgment with mere opinions.

When a teacher considers school ratings of any kind for the promotion or classification of a pupil, he is necessarily thinking of the achievement of the pupil in relation to his class group. Does he meet the standard in arithmetic? in geography? in deportment? in maturity, and the like? The problem of the teacher in this connection is to transmute the various achievements of the pupils, as they are revealed in the tests or otherwise, into letter, point-score, or percentile grades according to the system used in the schools where the teacher is employed. The various achievements

used in this determination should be of such nature that the teacher can be fairly certain of them in order to get a reliable grade. One of them is undoubtedly the standing of the pupil in the subject matter of the class, because so much school work is based upon that. Other elements have been suggested and should have weight according as the teacher has some objective standard by which to judge. The interest and attention of the pupils should be taken into consideration, but the teacher should remember that both interest and attention are later reflected in the degree of success attained by the pupil in the tests which are given. The maturity of the pupil is another consideration, but here the teacher should remember that the maturity of the pupil is reflected in the quality of the answers which he makes upon his tests. All in all, the safest ground for the teacher lies in the evidence for grading which is revealed in the results from Classroom Tests and Standard Tests. In exceptional instances these grades, where the teacher is sure of his ground, should be influenced by maturity, orderliness, and other facts of like nature.

The transmutation of scores into letter grades. A grading system is an attempt to separate a group into smaller units through distinguishing variations in the character of the work of the individuals who compose the group. This may be done in several ways, one of which is the M-scale system which has been described. This M-scale system, though exceptionally desirable in the preliminary steps of grading, may be unsatisfactory as a final mark because of the traditional systems in use where the groupings are of more general or of different character. These systems are commonly of two kinds: one in which the achievements of the pupils are classified on a basis of "good," "fair," "poor," and the like, and the other in which the pupils are given a percentile grade, sometimes on the basis of points converted into per cents.¹ Where these systems are in vogue, it is

¹ Compare discussion of percentile grades in Chapter X, pp. 201 ff.

difficult to use M-scale units as they stand, and it is wise for the teacher to convert these units into the prevailing system so that they may be understood. For convenience the first of these generally used systems will hereafter be called the letter-grade system, since letters are used to express the final form of the grades, and the second system, frequently used with it, may be called the percentile-grade system.

There are various forms of the letter-grade system now in use, the variations being in the number of groupings which are established by them and in the meanings which are attached to the groups. As a rule these meanings are fairly well established, as definitions at least, in the minds of the people who use them, and the reports as they are given out are fairly well understood. The following is a typical letter-grade system with five main groupings :

- A indicates excellent work.
- B indicates good work.
- C indicates fair, or average, work.
- D indicates passing work.
- E indicates unsatisfactory, or failing, work.

Another more or less typical system which is based upon the same number of groupings and upon much the same ideas, though using a different letter scale, is the following :

- E indicates "Excellent," or superior work.
- G indicates "Good," or work above the average.
- F indicates "Fair," or satisfactory work.
- P indicates "Passing," or poor work.
- U indicates "Unsatisfactory," or failing work.

Another scheme of the same sort gives figure designations to the groups instead of letters, thus :

- 1 indicates excellent work.
- 2 indicates good work.
- 3 indicates fair work.
- 4 indicates passing work.
- 5 indicates failure.

Such grades have the advantage of being generally used and are therefore quite generally understood, though an M-scale score or any other standard rating would be more satisfactory and if generally used would be as well understood. However, the problem of the teacher is to convert the scores made by the pupils into these letter grades, where such grades are used, and out of that problem grows the necessity for defining in some certain terms just what is meant by "excellent" or "satisfactory" or "poor." Any study of the marks given by teachers will show that whenever a school system has made no attempt to standardize the meanings of the ratings in any more definite form than they have been given above, the prevailing tendency is for the teacher to make such terminology mean whatever he will. A recent unpublished study of certain college markings, for example, showed that a certain teacher gave 30 per cent of his marks as A, 40 per cent of his marks as B, with the remaining marks distributed over the two other passing grades and with no failures. Another teacher in the same institution and with essentially the same group of students had over 15 per cent of his students classed as failures, 25 per cent of them classed in the lowest passing grade, about 40 per cent of them placed in the group called "fair," and the rest with grades of "good," with no students in the group of "excellent." Such a condition makes both the marks and their published meanings of little worth. An A with one teacher may mean a B or less with another. If a teacher carries on most of his work with one group of pupils, the standards used by that teacher will, of course, be used throughout the entire group of subjects which he is teaching, whereas variations in grading between teachers who work largely with different groups are difficult to detect on mere inspection. If two or more teachers have to do with the same pupils, however, the difference in conception as to the true meaning of the letter grades is more striking when the markings for the same groups of pupils are compared.

In order to make some headway in standardizing the grades within a class, or even within a school or larger educational unit, it is necessary to make some sort of agreement as to the marks as well as to the character of the pupils involved, unless, as will be shown, some definite standard of a better kind can be given. The real underlying situation is merely that if the marks are to mean anything at all they should indicate *relative* achievement, which in the classroom means the achievement of any one pupil in relation to the other members of the class. The first decision which the teacher must make is that with respect to the marking itself. Does the mark given mean the achievement of the pupil with relation to the rest of his class, or does it mean his achievement with relation to his possibilities to achieve? It is assumed here that there should be two such marks given to pupils: one showing his standing in the class and the other showing the degree of his effort. The discussion in the preceding chapter was devoted to the latter type of marking, whereas the discussion in this chapter is devoted to the former.

A second decision that the teacher must make is the meaning of the mark itself. Should a grade of A always mean the same achievement relative to the class group? This must be the case in order to be consistent with the first decision; and if a teacher agrees to the first decision without adhering to the second, the first has been merely restated and not accepted. In conformance with the second decision the teacher should make some definite effort to locate the exact range of each of the letter groupings. If, for example, a grade of B is always to mean the same thing, and if the letter B means "A certain range of achievement in relation to the class group," then the range of the class group or the percentage of the class group which should always receive the grade of B should be determined. In many places at the present time a way of determining this range of the group is to make some assumption, or proof if possible, as to the type of group which is to be marked, and to give grade groupings

according to that assumption. The most usual assumption is that the school group is in all probability an approximation to a normal distribution as far as ability is concerned; that there is no sharp break between different kinds of students or pupils; that the group, with respect to ability, has very few relatively poor individuals and about the same number of excellent individuals; and that the remainder of the group range themselves in gradually ascending and equally descending proportions about a central tendency which may be called the average. This we know to be practically true of the entire school population, and we know it to be less true with respect to small groups of pupils where many factors enter to disturb the regularity of the group. Moreover, we know the assumption to be more true of the lower than of the upper elementary school grades, because of the number of pupils who have dropped out before reaching the upper grades. In spite of the variations, however, and in spite of the objections which have been raised against this kind of grouping, it offers at least a suggestive solution to the problem of the meaning of the grades and one which if used will certainly tend to make the meaning of the grades more uniform and just. If such a system is not in use where one teaches, some scheme should be adopted which can be accepted pending a general school-system adoption. The scheme which the teacher adopts should be one that is as nearly as possible in accordance with his knowledge of his group. This knowledge can be reënforced objectively by means of intelligence tests, which, when carefully given, will give a distribution of a class group on the basis of what is commonly termed intelligence (more properly ability) and which can be taken as a basis for the grade groupings in the absence of a generally adopted system for school usage.

The following suggestive group percentages for grade grouping have been used in various places. In the absence of any adherence by a school system to such a scheme the teacher should choose the particular type of grouping which

best suits his needs. If none of those suggested is acceptable, it is possible for a teacher to devise a grouping of his own. The essential thing, in order to make the grades comparable with one another, is to be consistent in the use of whatever scheme is adopted. It will be noted that in the failure group there is always included a certain percentage of the pupils. The objection is frequently raised that "it seems too bad that someone always has to fail." The only answer to such an objection is that failure simply indicates an accomplishment considerably below that of the group, and is only called failure by inference. It is, in reality, not failure at all necessarily, but merely another division comparable to the division of B or C. Promotion is another matter entirely, and is the cause of the objection. It should be treated differently, as is brought out in a later section of this chapter. It is conceivable that we should, in elementary schools at least, do away with the term "failure" as such and think in positive terms rather than negative, — not how much does a pupil lack, but how much more does he need; not what has a pupil failed to do, but what can he do.

TABLE XXVII. PERCENTAGE GROUPINGS FOR FIVE-LETTER-GRADE SYSTEMS BASED UPON THE NORMAL DISTRIBUTION

LETTER GRADE	SYSTEMS							
	I	II	III	IV	V	VI	VII	VIII
A	3	4	5	6	7	8	10	15
B	23	24	24	24	24	24	24	22
C	48	44	42	40	38	36	32	26
D	23	24	24	24	24	24	24	22
E	3	4	5	6	7	8	10	15

In Table XXVII the percentage groupings are made upon the basis of the normal distribution and are calculated, given the A ratings, from the percentages cut off on the base line of such a curve in P. E. distances (P. E. (Probable Error) is a statistical measure of variability somewhat similar to the

S. D. that has been previously mentioned). In Table XXVIII the percentages are based upon minus skewed curves where it is assumed that there are more individuals of better quality than there are of poorer quality, which would result in more A's and B's than D's and E's.

TABLE XXVIII. DIFFERENT PERCENTAGE GROUPINGS BASED UPON ASYMMETRICAL CURVES, WHERE IT IS ASSUMED THAT THE GROUPS ARE MINUS SKEWED (COMPARE FIG. 11, CHAPTER XI)

LETTER GRADE	SYSTEMS							
	I	II	III	IV	V	VI	VII	VIII
A	3	4	5	6	7	8	10	15
B	22	23	24	25	26	29	31	32
C	54	52	50	48	46	42	38	32
D	20	19	18	17	16	15	14	13
E	1	2	3	4	5	6	7	8

To use these tables the teacher should merely decide which of the various percentage groupings is likely to be best in his situation and then adopt that for his own use. If none of those suggested seems applicable, any other may be constructed and used, though the systems given probably cover as wide a range of possibility as is necessary for most situations. Some teachers prefer to use percentage groupings more like those in Table XXVIII, where it is assumed that the group is skewed upward. If there are no school regulations, the system which fits best should be selected.

It may be found that the abrupt changing from the indeterminate system will arouse opposition of various kinds, and in some cases it has been found to be politic to change from one system to the other gradually. In one case, for instance, the mark of B had been consistently considered as the average mark, whereas a mark of C was considered passing, but approaching failure. The mark of D was rarely used and had almost no significance. The teacher felt that such a system was not in accord with the most desirable

practice, but at the same time wished, without arousing too great antagonism, to educate his pupils to a better meaning of A and B on a scale similar to those given above. His method, therefore, was to adopt a progressive scale system which from semester to semester gradually changed, cutting down the A and B groupings gradually until it was firmly established that a mark of C was not disgraceful but average, and a mark of A meant real distinction.

The more widely the grouping is used, as within various classrooms in the same grade, in all the rooms of a building, or in all the schools of a community, the greater does its value become, because it is more widely understood. Even if it is used by only one teacher, however, it is valuable, and when the teacher has made his choice of groupings he is then ready to use that scale in the making of his grades. The procedure is the same for any purpose which the teacher may have in mind, whether it be to give letter grades to a series of test papers or to make final grades for a semester or year.

STEP 1. RANK-ORDER DISTRIBUTION OF SCORES

The first step is to arrange the point scores, or M-scale scores, from which the ratings are to be derived, in a rank-order distribution. This can be done from the original raw scores when a single test is to be graded; but where a final grading is wished for a number of tests, the teacher should M-scale the raw scores for the separate tests and get a final M-scale composite which should then be arranged in rank order. It is not necessary to get a revised M-scale composite unless it is also to be used for an Accomplishment Ratio. It is unnecessary for the teacher to go back to the original undistributed scores if he has made a frequency surface in the way described in Chapter X, since the scores can be taken directly from that surface with less labor than by beginning again from the original scores.

JUDGING IN TERMS OF GROUP PLACEMENT 311

STEP 2. DETERMINING THE SCORE RANGE FOR PERCENTAGE GROUPS

The second step is merely to find the score range for each percentage group and to assign the letter grades according to that rating. In the example of test scores in Chapter X, p. 209, the distribution of scores is as given in Table XXIX. This is a rank-order distribution with the highest scores first.

TABLE XXIX. RANK-ORDER DISTRIBUTION OF SCORES ON TEST FOR ASSIGNMENT OF GRADES

61	48	43	40	38	35	32	29	23
51	46	42	39	37	34	32	27	20
49	44	40	38	37	34	30	25	15

Assume that the following letter-grade grouping has been adopted: A, 6%; B, 24%; C, 40%; D, 24%; E, 6%. It is then necessary to find out how many scores on the test that has been given constitute 6% of the total number of scores, for the A grouping, how many constitute 24%, for the B grouping, and so on for each of the letter-grade groupings listed above. Beginning with the highest scores (61, 51, etc.), the procedure is as follows:

The total number of cases is twenty-seven, and the computations for percentages of twenty-seven would appear as in Table XXX.

TABLE XXX. COMPUTATION OF PERCENTAGE OF SCORES TO BE INCLUDED WITHIN LETTER GROUPS

Grade	Percentage of Total	Number of Scores
A	6 % of 27	1.62
B	24 % of 27	6.48
C	40 % of 27	10.80
D	24 % of 27	6.48
E	6 % of 27	1.62

With this as a basis the teacher should adjust the markings, as far as the test itself is concerned, in discrete units.

One such adjusted group of markings might be as given in Table XXXI.

TABLE XXXI. THE CONVERSION OF TABLE XXX INTO DISCRETE UNITS

Grade to be given	Number of Scores
A	2
B	7
C	11
D	6
E	1

STEP 3. APPLICATION OF GRADES TO SCORES

The next step is to apply these grades to the actual scores. The two highest scores, 61 and 51, should be given

marks of A. The next higher seven scores should receive a grade of B. These are scores of from 40 to 49 inclusive. The next eleven lower scores should have a grade of C and would include scores of from 32 to 39 inclusive. The next six lower scores would receive grades of D and would include scores of from 20 to 30 inclusive. The single score of 15 would receive a grade of E.

It will be noted that if this plan were followed, one score of 40 would receive a grade of B and the other a grade of C. This should not be allowed to happen, and so a further readjustment is necessary. Both scores should be included in either the B group or the C group. In this case, since 40 is nearer the score of the C group than it is of the B group, it was decided to place both scores in the C group, reducing the B group to six scores and increasing the C group to twelve, as in Table XXXII.

TABLE XXXII. READJUSTMENT OF SCORES OF TABLE XXXI TO BRING ALL SCORES OF SAME SIZE INTO SAME GROUP

Grade to be given	Number of Scores
A	2
B	6
C	12
D	6
E	1

Then, by Step 3, the final grouping of scores would appear as in Table XXXIII.

It may well be questioned whether this is fair. It is not so fair nor so just as the actual M-scale scores would be, because of the much larger groupings and the assumptions as to the distribution of the class. It is much fairer, however, than letter-grade groupings unsupported by any standard whatever.

Transmutation of scores to percentile ratings. When the standard used in a school system is that of percentile groupings, such as those given in Table XXXIV, the procedure in reaching percentile scores from raw or M-scale scores is

TABLE XXXIII. GRADES AND SCORES AS DETERMINED BY STEP 3

Grades	Scores
A	61, 51
B	49, 48, 46, 44, 43, 42
C	40, 40, 39, 38, 38, 37, 37, 35, 35, 34, 32, 32
D	30, 29, 27, 25, 23, 20
E	15

TABLE XXXIV. DISTRIBUTION OF PERCENTILE RANGE BY LETTER GROUPS

Letter Group	Percentage Range
A	90-100
B	80-90
C	70-80
D	60-70
E	below 60

somewhat different. Let it be assumed that the percentile grading system involves a distribution by letter groups as shown in Table XXXIV.

STEP 1. DETERMINATION OF COMPARABLE SCORE RANGES FOR THE LETTER GROUPS

The first step in transmuting either raw or final M-scale scores into these percentage scores consists in the determination of the score range for each of the letter grades for both sets of scores. Where letter-grade groupings have not been made, the letter-grade distribution, as shown in the preceding sections, should be made. In the case cited above the comparative range is as indicated in Table XXXV, where it shows that the B group, for example, with a raw-score range of 42-51 has a percentile range of 80-90.

This gives a tabulated form which makes easier the further steps in the transmutation.

STEP 2. DETERMINATION OF RANGE DIFFERENCES

The next step is to determine the differences between the ranges, for both the raw scores and the percentile scores in

TABLE XXXV. RAW-SCORE AND PERCENTILE RANGES FOR LETTER GRADES

Grade	Raw-Score Range	Percentage Range
A	51-61	90-100
B	42-51 ¹	80-90
C	32-40	70-80
D	20-32	60-70
E	0-20	0-60

each of the letter groups. This gives for each letter group the total number of units in the range, through which comparisons may later be made.

In this case the difference between 61 and 51, the raw-score in the A group, is 10. The difference in the percentile range for the same group is 10. In Group B the two differences are 9 for the raw scores and 10 for the percentile scores ($51 - 42 = 9$, and $90 - 80 = 10$). Table XXXVI gives the complete tabulation of differences for the entire ranges of both types of score.

TABLE XXXVI. RANGE OF RAW-SCORE DIFFERENCES FOR DATA OF TABLE XXXV

Letter Grade	Raw-Score Range	Raw-Score Difference	Percentage Range	Percentage Difference
A	51-61	10	90-100	10
B	42-51	9	80-90	10
C	32-42	10	70-80	10
D	20-32	12	60-70	10
E	0-20	2	0-60	2

¹ Note that the upper limit of each score group is not necessarily the highest score in that group but is, in reality, the lowest score in the next higher group. The table should therefore be read, for example, "A grade of B has a raw-score range of from 42 *just to* 51 and a percentile range of from 80 *just to* 90." Perhaps the table might be more accurately written as 42-50.99999 and 80-89.99999.

² Because of the great differences between the zero points it is advisable to calculate the differences for the E letter grade. See the later section on "Special consideration of E-score groupings."

STEP 3. DIVISION OF PERCENTILE DIFFERENCES BY RAW-SCORE DIFFERENCES

The third step in the transmutation is to divide the percentile-scale range differences by the raw-score range differences for each letter group. This gives the number of percentile-scale units which are equivalent to each raw-score unit for the letter grades in question.

In Group A the percentile difference, 10, should be divided by the raw-score difference, 10, which would give a result of 1.00. This means that for each raw-score unit in Group A there is 1.00 unit in the percentile scale. In Group B, by the same procedure, there are 1.11 percentile units for each raw-score unit. Table XXXVII shows the results as computed in this way for all the cases above cited. In each case the percentile difference is divided by the raw-score difference.

TABLE XXXVII. NUMBER OF PERCENTILE UNITS FOR EACH RAW-SCORE UNIT, BY STEP 3

Letter Grade	Raw-Score Difference	Percentile Difference	Proportion of Raw Score to Percentile Differences
A	10	10	1.00
B	9	10	1.11
C	10	10	1.00
D	12	10	0.83
E	— ¹	— ¹	0.83

From this point on in the transmutation of raw or M-scale scores to percentile scores there are two methods, both of which yield the same results, for use according to the number of cases in the various letter groups. When there are only a few cases, or when the cases are in uneven steps in a group, the first method given below is preferable because it takes less time in its computation. When, however, there are a number of different cases, the second method will be found preferable.

¹ See footnote 2, p. 314.

METHOD TO BE USED WHEN FEW CASES OCCUR IN ANY SCORE GROUP

STEP 4. DETERMINE SCORE DISTANCE ABOVE LOWEST SCORE

The next step is to determine for each of the raw scores the difference between that score and the lowest raw score in that group. This gives the degree to which each score is above the lowest score of the group in terms of raw-score units.

TABLE XXXVIII. UNIT DISTANCES ABOVE LOWEST SCORE FOR B GROUP

Raw Score	Unit Distance above Lowest Score
49	7
48	6
46	4
44	2
43	1
42	0

The A group in the case mentioned above contains only two cases, one of which is the lower and the other the upper limit of the group; so these may be directly transmuted without further computation into 61 as 100 per cent and 51 as 90 per cent. The B-score group can be used as an illustration of this step, however, as it is about as efficient as the method later to be described. The six cases in the B-score group of the raw scores and their unit distances from the lowest score in the group are as given in Table XXXVIII, where each unit distance is calculated from 42.

STEP 5. MULTIPLICATION OF UNIT DISTANCES BY PROPORTION FOUND IN STEP 3

The fifth step is to multiply the differences thus found between each score and the lowest raw score by the amount found for that score group in Step 3. For the B-score group this was found to be 1.11 (see Table XXXVII). If the teacher wishes, these figures can be taken to the nearest first decimal, which in this case would be 1.1. Although this is somewhat shorter, it is not so accurate. The tabulation of this computation is as given in Table XXXIX.

STEP 6. ADDITION TO LOWEST POINT OF PERCENTILE RANGE

The final step in the transmutation of these scores is to add the results reached in Step 5 to the lowest point of the percentile scale for that letter-grade group. In the case listed above, the lowest point in the percentile group, as shown in Table XXXIV, p. 313, is 80 per cent. A score of 42 is therefore equal to 80.00 plus 0.00 per cent, or 80.00 per cent.

A score of 46 is equal to a per cent of 80 plus 4.44, or 84.44 per cent. For this B group of scores the transmutations will stand as in Table XL. The first column gives the raw scores; the next column shows the lowest percentile point for the group; the third column shows the amount to be added as derived in Step 5, Table XXXIX, and the last column gives the final transmuted percentages for the group.

A sheet of paper arranged in the tabular form of Table XL, and containing space for the other score groupings, A, C, D, and E, will be found to be a great help in keeping the calculations accurately.

TABLE XXXIX. MULTIPLICATION FOR GROUP B BY PROPORTION DERIVED IN TABLE XXXVII

Raw Score	Unit Distance above Lowest Score	Multiplied by 1.11
49	7	7.77
48	6	6.66
46	4	4.44
44	2	2.22
43	1	1.11
42	0	0.00

TABLE XL. ADDITIONS OF AMOUNTS FOUND IN TABLE XXXIX TO LOWEST PERCENTILE POINT IN GROUP

Raw Score	Lowest Percentage Score for Group	Amount to be Added (Table XXXIX)	Final Percentage
49	80.00	7.77	87.77
48	80.00	6.66	86.66
46	80.00	4.44	84.44
44	80.00	2.22	82.22
43	80.00	1.11	81.11
42	80.00	0.00	80.00

METHOD TO BE USED WHERE THERE ARE MANY CASES IN THE SCORE GROUP

STEP 7. SUCCESSIVE ADDITION TO LOWEST PERCENTILE SCORE

Instead of using the foregoing steps when the total number of cases in any raw-score letter group is as great as the number of possibilities, or nearly as great, the following procedure may be substituted with some saving of time:

TABLE XLI. SUCCESSIVE ADDITIONS IN B GROUP TO REACH FINAL TRANSMUTED PERCENTAGES

(Add and read up)

Score	Amount Added	Percentage
51	1.12	90.00
50	1.11	88.88
49	1.11	87.77
48	1.11	86.66
47	1.11	85.55
46	1.11	84.44
45	1.11	83.33
44	1.11	82.22
43	1.11	81.11
42	{ Score assigned }	80.00

First *all the possible raw scores* for the group, whether some have been actually received or not, should be listed without breaks from the lowest to the highest, as in Table XLI, and the lowest score for the percentile group should be assigned to the lowest raw score. Then the number found for that group in Step 3, p. 315, should be added successively to the first percentile score to reach the final scores for the group.

In the case just cited the scores would have been listed as in Table XLI, making sure that there were no gaps in the series of scores, even though there might be no actual scores for some of the cases. The table has been arranged to make it read in the same way as the other tables given above.

It may be seen that this method reaches the same percentage ratings for the raw scores as does the previous method described. It will also be noted that in the final addition an extra hundredth has to be added to bring the percentage numerically even with the previously assigned percentage for the score of 51. This is occasioned by the

fact that the number used for additions, 1.11, is a little less than the true number, 1.111

Summary of steps in transmutation of raw scores to percentage ratings. For convenience the steps above outlined are here repeated in condensed form :

1. Determine point ranges for each letter grade for both raw and percentile scores.

2. Divide differences between point ranges for both raw and percentile scores for each letter group.

3. Divide, for each group, the difference found in the percentile range by the difference found in the raw or M-scale score range.

Use as follows with few scores in a group :

4. Determine for each raw score the difference between that score and the lowest raw score in that group.

5. Multiply that difference by the amount found in Step 3.

6. Add this result to the lowest point in the group range of percentile scores, which gives the final desired transmuted percentage mark or grade.

Or use this when there are many cases in a group :

7. Add successively to the lowest percentile score in each letter group the amount found in Step 3 for that letter group. Each successive addition will give the percentage score for each successive higher raw or M-scale score.

Special consideration of E-score groupings. The E-score grouping in both the raw-score units and the percentile units includes all the scores below the lowest score in the D group. Since the zero point of the percentile scale is so far below the zero point of the raw-score scale, it is useless to attempt to carry out this procedure in that group and it is better to subtract successively from the lowest D percentile score for each raw-score point below the lowest D-group score, in order to find the few E percentage ratings. In the foregoing example the interval for each of the raw scores on the percentile scale for the D group (see Table XXXVII) is 0.83. The single E-group score is five raw-score units below

the lowest D-group score. Multiply the intervals of the D group, 0.83, by 5, and *subtract* this from the lowest D percentile score, 60.00, to find the percentile grade for the score of 15 ($0.83 \times 5 = 4.15$; $60.00 - 4.15 = 55.85$). Another way is to follow the second suggestion given in the directions in previous sections, and *subtract* successively 0.83 five times from 60.00, which would give the same percentage, 55.85.

Completed transmuted raw scores to percentile grades. Table XLII shows the completed transmutations for all scores derived in the manner shown above.

TABLE XLII. COMPLETED TRANSMUTATION TABLE

Raw Score	Percentage	Raw Score	Percentage
61	100.00	37	75.00
51	90.00	35	73.00
49	87.77	34	72.00
48	86.66	34	72.00
46	84.44	32	70.00
44	82.22	32	70.00
43	81.11	30	68.30
42	80.00	29	67.47
40	78.00	27	65.81
40	78.00	25	64.15
39	77.00	23	62.49
38	76.00	20	60.00
38	76.00	15	55.85
37	75.00		

The use of tests for promotion. Promotion of pupils depends upon several factors, one of which is, of course, the standing of the pupil in his own group. A more important consideration in many respects is his ability to do the work of the grade or class to which he might be promoted. Another important consideration is his social and physical maturity, although as far as elementary schools are concerned this is of less importance between grades than it is between the elementary school and the high school. The two most important considerations are the first two men-

tioned; and of the two, the first, the standing of the pupil within his own group (the evidence of his past achievement with respect to that group), has in the past been the more generally used criterion for promotion both in school and in college. The question of the pupil is ever, "Did I pass?" He assumes that if he did pass he is qualified to undertake the next higher curricular step in its entirety. The second criterion for promotion, the ability of the pupil to do the work of the grade to which he may be promoted, has been made possible only through the use of Standard Tests. It is less generally used than the first criterion, but is much superior to it, and its use should increase as time goes on. Where a group of Standard Tests is given, each pupil, as far as those tests cover the work he has done, is measured with respect to his past achievement, and the point of his next effort is accurately placed. If that *next effort* lies in the next upper grade, he should be promoted; but if it does not lie there, he should not be promoted. This should be the criterion for measuring success and failure. Since the Standard Tests are standardized for both grade and age or are convertible one into the other, it is easily possible to test whether or not a pupil is capable of doing the work of the next higher grade and whether or not he is ready to do it.

As was stated, however, the first criterion is at present more generally used than the second, and in addition it has depended largely on the traditional forms of teacher's examinations, the disadvantages of which in the elementary school, were discussed in an earlier chapter. The tests described in this book will give to the teacher using this criterion a better basis for making promotion than these traditional examinations, though it would certainly be unwise for a teacher to depend upon them alone for his entire judgment of the pupil. The Classroom Tests will give an indication of the results which the pupil has accomplished and will also show to the teacher the pupils at the lower end of the class, whose promotion would be doubtful; but the

teacher should also include in his final judgments his own knowledge of the pupils. This knowledge of the pupils is undoubtedly subjective, but it is nevertheless valuable. A judgment as to the initiative or the attention of the pupil, as to his physical and social maturity for the work of the next higher grade, as to his steadfastness of purpose, as to his tendencies to coöperate with his teacher and the other members of his class, and as to his habits of study is certainly important in estimating his worthiness for promotion and should be taken into consideration along with the results of the tests to which he has been subjected.

The tests that have been described will show where the pupil stands with relation to his work and with relation to his class group. As such they are valuable in reënforcing the judgment of the teacher or even in correcting it. There can be, however, no set rule for promotion on the basis of these tests, any more than on the basis of the traditional school examinations, and the prevailing rule in effect in a teacher's community should be followed. The teacher should, moreover, appreciate the fact that his judgment of the worth of the pupils is considerably improved as a result of using these tests.

When used for promotional purposes in combination with the Standard Tests, where provision is made for the inclusion of a teacher's judgment, the results of these tests may furnish an excellent means of confirming or correcting that judgment, and in this respect the judgment of the teacher is on practically the same objective basis as that of the Standard Tests themselves.¹

The use of these tests for classification purposes. There are two bases at least upon which classification of pupils can be made as a result of giving these tests, and upon either basis there are three types of classification with which a teacher has to deal or ought to deal, all of which can be materially advanced and in most cases entirely completed through the

¹ Compare W. A. McCall, *op. cit.* chap. ii.

medium of these tests. The two bases of classification are classification in terms of Achievement Ratios and classification according to placement within the class group. The teacher can use the first type of classification very advantageously in classifying on the basis of effort the pupils taking certain subjects. The second type of classification can be used for classifying pupils on the basis of actual achievement. The three types of classification which can be used with either of the above two bases are classification within a single schoolroom, classification within a single grade, and classification within an entire school.

Classification within a room. Classification within a room is probably the most prevalent type with which the teacher has to deal. There are times when he wishes to have two or more groups of pupils with about equal abilities. These may be called equivalent groups. There are other occasions when the teacher may wish to have one half of the class, the more capable in certain fields, doing special types of independent work, and the other half of the class, the less capable in that particular phase of the work, concentrating on work designed to meet their needs. These are complementary groups. At other times the teacher might wish to have a small group set apart from the rest for special purposes. The results of the tests described in this book should be of great help in determining the composition of these groups.

Equivalent groups. To divide the pupils of a classroom into two equivalent groups on the basis of these tests, the test results for the entire group should be tabulated in rank order from best to poorest. This can be done directly from the raw scores when only one test is used as a basis, but the classification into equivalent groups is of course better with the increased number of tests as a basis for division; and therefore, when more than one test is used, the M-scale composite should be found. When the class group has been arranged in rank order according to the tests, the group can

be separated by taking the first pupil for Group 1, the next two pupils for Group 2, the next two pupils for Group 1, the next two for Group 2, and so on for the separation of the pupils in the entire class. The result will be two groups of pupils of approximately equal achievements. When Achievement Ratios are used, the system is the same, and the result will be two groups of pupils equivalent in their effort.

Table XLIII divides into two equivalent groups, on the basis of raw scores, the group shown in the distribution in

TABLE XLIII. DISTRIBUTION OF PUPILS INTO TWO EQUIVALENT GROUPS

Score of Pupil	Group	Score of Pupil	Group
61	1	37	2
51	2	35	1
49	2	34	1
48	1	34	2
46	1	32	2
44	2	32	1
43	2	30	1
42	1	29	2
40	1	27	2
40	2	25	1
39	2	23	1
38	1	20	2
38	1	15	2
37	2		

Chapter X, p. 211. The first column shows the raw scores of the pupils, and the second column shows the numbers of the groups to which the pupils reaching the various scores should be assigned. M scores can be used in exactly the same way.

Complementary groups. Complementary groups are found in much the same way as are equivalent groups. After the pupils' scores have been arranged in rank order from best to poorest, the group should be divided according to the purpose the teacher has in mind. If two complementary groups are desired of approximately equal size, the rank distribution

should be divided in halves, the upper half forming one group and the lower half forming the other. If three groups are desired, the group should be divided into thirds, the upper third forming one group, the lower third forming another, and the remaining middle third constituting the complementary connection between them. Table XLIV, made from the data used above, shows three complementary groupings of pupils as described.

TABLE XLIV. CLASS DIVIDED INTO THREE COMPLEMENTARY GROUPS

Grouping in complementary groups according to Accomplishment Ratios is also easy. If two groups are wanted, those whose A. R.'s are at or above 100 could form one group, and those with A. R.'s below 100 could form a second. If three groups are desired, those above 100 could form one group, those at or near 100 a second, and those below 100 a third.

SCORES		
Group 1	Group 2	Group 3
61	40	32
51	39	32
49	38	30
48	38	29
46	37	27
44	37	25
43	35	23
42	34	20
40	34	15

Special groups. Special groupings of the better pupils for additional forms of work, or small groups of the poorer pupils for increased amounts of drill or for any other purposes, may be determined after an analysis of the test results and can be selected from the remainder of the class upon that basis. The size and quality of the group and the character of the pupils would of course depend upon the wisdom of the teacher, reënforced by an analysis of the test results.

Classification within a grade. Classification of all the pupils within a grade into two, three, or more groups can also be satisfactorily accomplished by the use of test results, either from raw and M-scale scores or from Accomplishment Ratios, although in this case the same tests or series of tests should be given to all the pupils simultaneously and the results should be treated as though they were all from one

classroom. After the results have been tabulated as is shown in preceding chapters, they can be used for the various classifications which were discussed above, and in exactly the same way.

Equivalent groups. For equivalent groupings, that is, two or three divisions of pupils of approximately the same ability, the scores should be arranged in rank-order distribution, best score to poorest score. For two groups the selections should be made as in Table XLIII. When three groups are desired, every third pupil should be chosen for the first group, of the remainder every other pupil should be chosen for a second group, and the remaining pupils should constitute the third group.

A somewhat fairer scheme is to choose the pupils on the following sequential basis :

Pupil 1	Group 1	Pupil 5	Group 2
Pupil 2	Group 2	Pupil 6	Group 1
Pupil 3	Group 3	Pupil 7	Group 1
Pupil 4	Group 3		Etc.

The first six elements of the sequence should be repeated in the selection until all the pupils in the grade have been assigned to groups.

Complementary groups. When complementary groups are wanted, the entire group of pupils should be arranged in their rank order and then should be divided according to the number of groups desired. If two groups are wanted, the rank distribution should be divided into halves; if three groups are wanted, the rank distribution should be divided into thirds; and so on for other numbers of groupings.

Special groups. Special groupings can also be selected from the entire group on the same basis as has previously been described, and where there is departmentalized work in the lower grades the use of these tests will prove of great aid in making the proper divisions of classes, according to the desires of the teacher or to the needs of the pupils.

Classification within a school. Classroom Tests, limited as they are to grade and particularly to individual-classroom use, are of only limited value in the classification or reclassification of a school. What is needed in this case is a single group of tests which can be given to practically all the pupils in a school and the results of which can be compared with uniform age and grade standards. This is a field for the use of Standard Tests, and they can be used in batteries of many tests or singly, as preferred. It is only when the method used in the reclassification takes into account the judgments of the teachers of the various classrooms that these tests can be used to advantage. When so used they serve to reënforce or correct the personal judgment of the teacher, as has been mentioned.

Chapter summary. Teacher's Classroom Tests may be made of real aid to teachers in marking or grading the work of pupils, in promoting them to higher grades, or in classifying them in groups.

In marking or grading, the raw scores obtained on the tests, or the M-scale scores which result from the composite of a number of tests, may be used when it is necessary to convert them into a prevailing letter-grade or percentile grade system. In either case it is merely a matter of transmuting these test scores into grades or percentile marks according to the system outlined in the chapter.

In promotion the tests can be used advantageously for the purpose of reënforcing or correcting the judgment of the teacher, although it will be found wise as a rule for the teacher to take other matters into account in the recommendations for promotion, because the tests do not and cannot test the wide range of qualities and knowledge which should be implied in promotion.

For classification purposes the tests can be used alone where achievement classifications are wanted, or the Achievement Ratios can be used where effort classifications are wanted. The classification may be in terms of equivalent,

complementary, or special groupings, and either in a single classroom or, if the tests have been used so widely, in a single grade. Where classification is necessary outside of these limits, Standard Tests should be used, and Classroom Tests can become only a means of influencing the judgment of the teacher.

SELECTED BIBLIOGRAPHY

- RUGG, H. O. "Teacher's Marks and Marking Systems," in *Journal of Educational Administration and Supervision*, February, 1915.
- CUBBERLEY, E. P. *The Principal and his School* (chap. xix, "Classifying and Promoting Pupils," and chap. xxiv, "Measuring the Instruction"). Houghton Mifflin Company, Boston, 1923.
- STEBBINS, R. C. "The Accomplishment Quotient¹ as a Means of Classification in the Lower Grades," pp. 34-44 of *First Year Book*, Department of Elementary School Principals. National Education Association, 1922.
- TRABUE, M. R. *Measuring Results in Education*, pp. 482-485. American Book Company, New York, 1924.
- MCCALL, W. A. *How to Experiment in Education*, pp. 45 ff. The Macmillan Company, 1923.

¹ The Accomplishment Ratio is the same as the Accomplishment Quotient but is more recent terminology and indicates a ratio to mental ability. The term "quotient" is being restricted to ratios indicating a relation to chronological-age level.

CHAPTER XVII

SPECIFIC USES: THE USE OF CLASSROOM TESTS AS DEVICES IN TEACHING

Classroom Tests as a teaching-device. An examination or a test is largely the result of the necessity or desire of a teacher to measure his pupils either in respect to the extent of their individual effort or in respect to the extent of their individual achievement. It results in what has been previously described as diagnosis, improvement of teaching, the finding of Accomplishment Ratios, the classification of pupils, and the like. A teaching-device is the result of the necessity or desire of a teacher to find materials, not so much to measure what has been taught but rather to facilitate the teaching itself. It results in definite provisions for the use of the Thorndike "Laws of Learning" in all their aspects: first, in devices to help pupils to become ready to learn; secondly, in devices to exercise the elements of that learning; lastly, in devices calculated to bring home to the pupil satisfaction in right learnings and annoyance at wrong learnings.

The Classroom Tests, to which this book is devoted, if judiciously used can help as teaching-devices to promote these ends. Largely for this reason the various tests throughout the book have been called papers in an effort to minimize, if possible, the testing phase of their results in order to emphasize so far as might be the devices phase.

Teaching-devices may, however, have either a desirable or an undesirable (at least questionable) set of characteristics. So far as readiness is concerned, the distinction as to whether the devices used are desirable or undesirable is largely according to the character of the devices themselves

and the way they are used. Readiness varies in great measure as do its concomitants — interest and attention. If any device divides the interest or attention with itself, it may be considered undesirable. This is what is meant by "sugar-coating," a device for making learning attractive through hiding the essentials of the learning behind something instinctively or by learning more attractive to pupils than the elements hidden. It makes the elements through which learning takes place mere adjuncts of the coating, and under such conditions, although pupils may seem to be learning, they may merely be absorbing the sugar coating. The fundamental difficulty here is the fact that the interest and attention, from which the readiness results, are external to and not inherent in or a part of the elements from which learning results.

The desirable type of device is that which merges itself (the interest and the attention which is paid to *it*) with the elements which are being used in the learning process. Thus, through the device, the readiness to do the work in hand is increased and the value of the work, as well as of the device, is enhanced. In this type of device the interest and attention are not external to the learnings which result but are a part of them.

It has been the experience of many teachers who have used these tests that they may easily be made desirable devices. Used so that pupils do not become wearied with testing, and including really thought-provoking elements that can be discussed, the tests are welcomed by pupils. More than that, the learning which the papers help to bring about is as eagerly welcomed. The papers supply objectives for study, and motivation for discussion of a new type and of a valuable character.

With respect to exercise the distinguishing element which makes the exercise valuable is whether or not the pupil actually wants to improve. If the pupil wishes to improve, devices which bring about exercise will undoubtedly result

in improvement, provided the pupil is told when he is right and why he is wrong. The spelling-bee, sets of exercises in arithmetic, recitations, or what not serve to give desirable exercise to pupils who wish to improve and mean but little to pupils who do not. In general it is probable that these exercises help the more capable (those who *can* and *do* win) more than they do the less capable.

A device which can reduce the element of exercise from rivalry of others, where only a few can win and where the same few usually do win, to rivalry of one's best previous efforts, where all can win and have an equal chance to win, cannot help being desirable. When this device, in addition, contains within itself the possibility of pointing out the elements of rightness and wrongness, it becomes even more valuable. Classroom Tests by themselves promote some desirable exercise, but when used in conjunction with Achievement Ratios and careful reviews they are a splendid and worthwhile device for this purpose.

The effects of the exercise, — satisfaction and annoyance, — may also be desirable or undesirable. It depends on whether satisfaction is attached to those learnings that should be encouraged and annoyance to those which should not be encouraged. A further element of the law of effect might be whether or not satisfaction was attached to the whole process, so that readiness (the beginning of the chain of learning) might result.

There can be little question that Classroom Tests, from a general point of view, when the scores are protected from divulgence as has been suggested in previous chapters and when the achievements of pupils are measured according to their several abilities, are especially well adapted as a teaching-device to bring the law of effect to bear. From the specific point of view Classroom Tests are well adapted to have satisfaction or annoyance attached to their elements, because when each specific element is explained and discussed after a test the proofs of success or error are present. The result is that

instead of the pupil's attaching his annoyance over errors to his teacher or to schooling, or to any extraneous thing or circumstance, he directly attaches it to his own error, where it will do the most good. Moreover, because successes are also specific, it means that there is no favoritism or prejudice, and successes give real and not doubtful satisfaction. In order to bring this about in most complete fashion the teacher should hold himself ready at all times, unless he has good objective reason for not doing so, to reverse or change his opinions or judgments when he believes that signal errors in scoring and the like have been committed.

In addition to testing, then, Classroom Tests are excellent devices which can be used and have been used by classroom teachers. They exhibit the desirable aspects of such a classroom device by making pupils interested in learning for itself, by offering a type of exercise in which desirable activities are promoted, and by making it possible both, in general, through Achievement Ratios and, in particular, through the reviews which bring out and specifically correct the errors that have been made.

Use of tests as a teaching-device in upper grades. If precautions are taken to insure as complete an understanding as possible of the tests which are given, and especially to have the results of the tests specifically connected with the work of the class, it should not be difficult for teachers to use Classroom Tests in the upper grades as a teaching-device. Some of the procedures which have been found to contribute to success in this connection are enumerated below :

1. *Scoring and checking by pupils.* In spite of the fact that, as a rule, scoring of papers by pupils contains more possibility for errors than does scoring by the teacher, it is nevertheless a valuable procedure. It gives in the first place an opportunity, which is otherwise impossible, for discussion of certain points by the class. This after-test discussion is especially valuable, since it comes at a time when the problems involved in the papers are still fresh in the minds of the

pupils and when the *reasons* which impelled the pupils to make their answers can still be recalled. In the second place it is valuable because it enables the teacher to put emphasis upon principles which appear from the class discussion to have been somewhat misunderstood, an emphasis which has more force and more reason behind it than it can have at any other time. In the third place it is valuable because it allows the teacher to see his own errors in the test, such as catch questions previously unobserved, ambiguous statements the ambiguity of which had not been apparent, and the like, and to make such adjustments at the time as seem called for. In the fourth place it provides for a discussion of the relative values of different types of answers and for class decisions on a socializing and coöperating basis as to the values which should be given. It gives the teacher a splendid opportunity to become a part of a class.

The chief difficulties in scoring by pupils, as has been stated in previous chapters, are the greater opportunity for error and the greater time required. By the general method given below, the first difficulty can be removed or minimized, and of the second objection it need only be said that it is doubtful whether or not this extra amount of time could be devoted to any better use.

This general method of scoring by pupils is as follows: When the papers have been completed, they should be collected and passed out again in different order from that in which they were taken up. If they were collected from the front and were redistributed from the rear, the central group of pupils would be very likely to get back their own papers. A way to avoid this is to have the papers passed up to the front of the room, collected, and then redistributed by starting about two thirds of the way back in the room. This effectually mixes the papers so that no pupil is likely to receive his own.

It is a good plan to have the pupil who corrects a paper mark it with his name or initials following the words "Cor-

rected by." This will serve both to prevent many errors of careless marking and to locate the responsibility for any errors that may be made. Under these conditions pupils are attentive to discussions and scoring decisions, and are quick in defense of their own answers if they believe them to be well grounded. The teacher will usually find that class opinion, save in a very few cases of widespread misunderstanding, will have a more salutary effect than any teacher's decision could possibly have and will carry just as great weight in bringing proofs of error to pupils. He will also find that he can be more of a leader and can exert that leadership with less dogmatism under these conditions.

By re-checking papers occasionally and by carefully considering such objections as may be brought up, it is possible for the teacher to be assured of fairly accurate results. The teacher should remember, however, that the larger a class is the more difficult it becomes to keep out errors in scoring under these conditions, especially where interpretations must be made of answers, as in the Judgment Test. All in all, the advantages in scoring by the pupils far outweigh the disadvantages in their effect upon the class from the point of view of teaching. When diagnosis and improvement of teaching are the prime motives of the testing, it is probable that the teacher should score the papers himself, since in that way he is enabled to gain that larger view of class achievement which is necessary.

2. *Graphic presentation of results.* Graphic presentation is always more emphatic than mere oral presentation of the results of a test, and it serves to let pupils see exactly what they have done with relation to the rest of the class. One of the test pictures that can be made is the frequency surface of the scores. Here the use of colored chalk to designate the various parts will be of help in interpretations. The teacher should designate the score groups below the base line and should place the exact scores in squares above the base line. He should *not* place any identification mark in any of the

squares, but should teach each pupil to find out for himself where he stands with relation to the rest of the group.

Probably a more generally useful graph for pupils is that of the question difficulty, a picture which will show each pupil the relation of his errors to those of the class in general. In this graph the question numbers should be clearly written, so that the pupils can identify the various questions for comparison with their own papers.

Whenever such graphs are used they should be carefully explained, especially until the pupils are familiar enough with the meanings to be able to make their own interpretations. The use of such graphs, however, should help in making the tests a real teaching-device and should be of actual service to pupils as well as to the teacher in locating objectives.

3. *Self-rating record cards.* Perhaps one of the more valuable of the procedures that are possible in helping these tests to contribute to teaching — more valuable because of the type of motivation which it fosters and more valuable because of the interest which it engenders — is the use of self-rating achievement cards. The card suggested in Chapter XV, p. 298, is one of this type. With this a pupil can record his progress, and through it he can note his improvement. If he makes his own Achievement Ratios, places them on his card himself, draws his "line of improvement," and judges his own progress in relation to those ratings, he has gone a long step toward finding good and useful reasons for doing his best work. Under such circumstances a teacher will be occupied more in teaching pupils how to learn to better advantage than in finding methods by which to prod them. He can work positively rather than negatively; he can lead his pupils rather than drive them.

The use of Classroom Tests as a teaching-device in lower grades. In the lower grades, especially in the primary grades, probably the greatest value of these tests, in the writer's opinion, lies in their use as a teaching-device. The tests have been most satisfactory wherever they have been used in the

first, second, and third grades, but their use, under the writer's observation, in the first and second grades has been too limited to enable him to form a worthy judgment as to their value. The few teachers who have used the tests have been very ingenious in adapting the test types, in content, purpose, and administration to the abilities of their pupils and have done highly suggestive work. The papers can be dictated, written on the blackboard, typewritten, written by hand on stencils for mimeographing (which presents the materials in familiar script), or typewritten on stencils as previously described. Any way that reduces the mechanical difficulties for little children will be satisfactory.

In the following examples the value of the tests as a teaching-device for comprehension and word knowledge should be plain. The tests formed part of a battery given by Miss Alice M. Brennan, of Cherry School, Toledo, to first-grade pupils early in the second half-year. Though the tests seem difficult, the test analyses showed them to have been well adapted to the abilities of the pupils.

Miss Brennan shows her method of giving the tests as follows:

TEST A. SELECTION BY ELIMINATION

Each pupil was given a mimeographed sheet, face down. The pupils wrote their names on the backs of the sheets, the following instructions were given, and then the papers were turned over for work.

Instructions: "Read the words on this paper. The words on each line make a sentence. There is one word that does not belong with the other words. Find that word and cross it out. Read your sentence again and see if it makes good sense without the word you crossed out."

TEST B. COMPLETION BY SELECTION

The sentences were given in mimeographed form with the directions there included.

Sample Papers of a First Grade

FIRST-GRADE READING PAPER

SCORE_____

DATE_____

[Directions as given above]

1. cows corn like pretty
2. the green grass fox is
3. sing birds see songs
4. cat was saw a mouse the
5. fox the after hen ran a with
6. Red Hen wheat will plant tell
7. Mother Goat seven had green kids
8. goats over went the bridge he a
9. apples seeds have in blue them
10. meadow sheep tall are the in the
11. nest in tall is the tell tree a
12. Peter Rabbit from away ran mother his her
13. we eat should bread tea
14. children across run well must not the street
15. help we each other eat

FIRST-GRADE READING PAPER

SCORE_____

DATE_____

Here are some sentences. A word is missing from each one. Look in the words at the side and find the word you think belongs in the sentence. Draw a line around the right word.

1. Birds — nests in trees. help bring build eggs
2. Flowers grow in a —. girl garden green gold
3. Children like to play with —. tell two talk toys
4. We slide down a —. brook hall house hill
5. The father robin's head is —. red brown black orange
6. The robin's eggs are —. black blew blue blow
7. Children should be kind to —. another animals toys
they
8. We should drink —. much milk coffee look
9. A mother is called —. Mrs. Miss Mr. my
10. The baby birds are fed by the —. boys feather father fat

Another example of a Selection Test adapted for use in a second grade is given below. The test was constructed by Miss Edna Spilker, Newton School, Toledo, and was given by typing or mimeographing the two sets of elements on two different colors of paper, yellow and white. The elements were cut apart, and each pupil was given a complete set of the test, with directions to match the yellow parts with the white parts to make good sentences. The ability to be tested was that of sentence structure, and though perhaps of the nature of a test it was certainly far more of a teaching-device.

Sample Second-Grade Paper

SERIES I

When the white men came to America
The red men were
Our land was then covered
The Pilgrims made their houses
Little Indian boys learned
The Indian men spent much time
The Pilgrims asked the Indians
The Indians made
The Pilgrims thanked God

SERIES II

to shoot with bows and arrows.
"Will you come to our feast?"
beautiful baskets from grasses.
they found red men living here.
for keeping them safe and for their good harvest.
called Indians.
from the skins of animals.
hunting wild animals.
from logs which they cut in the woods.
with thick forests.

Conclusion. In conclusion, for the teacher who uses this book there are a few advisory precautions which may be of help, first, to make any work which is done less tedious and,

secondly, to make the results more useful. These can be listed as general precepts for the successful use of Classroom Tests.

1. *Use tests for specific purposes.* If the teacher knows neither what he wishes to test nor why, it is probable that both he and his pupils will be better without tests. It is always a good plan to decide first why one is going to test, and then to build the test about that as a basis.

2. *Plan to derive only those results that are in terms of the purposes.* Any results that can be derived from Classroom Tests are likely to be of interest, but they are not all needed with every test. If Accomplishment Ratios alone are wanted, it is not necessary to make the frequency surfaces or graphs of question difficulty. Frequently such diagnosis as is needed can be made sufficiently well for temporary purposes from judicious inspection. The teacher's time is short, the teacher's energy is limited, and there are only twenty-four hours in a day. The teacher should plan for the results that are wanted and stick to that plan.

3. *Plan to teach and test in terms of principles rather than mere facts.* "The world is so full of a number of things" that we can hardly expect our pupils to know them all. Principles, on the other hand, have universal significance and universal application. It makes school far more worth while to teach principles through facts and to test principles rather than the knowledge of facts.

4. *Attempt the new after the old is mastered.* It has been hoped to have this book show a reasonable progression of processes. It is not expected that a teacher shall learn all the processes at once; it is not hoped that he will attempt to. Each unit in itself has a certain value, but each succeeding unit gains value from those that have come before. Therefore the processes in each unit can be mastered separately, and each should be mastered before others are attempted. The teacher should remember the fable of the father and his boys with the bundle of sticks. They could not break the entire bundle at once, but they *could* break

each stick separately. First, the teacher should try to make the tests. When he has made good tests, he should follow with any processes that appeal to him and master each in turn. He will shortly find that each process is simple and that the results are good. Discouragement can come only from attempting to do too much or from doing a reasonable amount too quickly.

5. *Keep records carefully.* The teacher will find that order in processes and neatness in work are splendid assets. Much time will be saved if a teacher will carefully provide for each step that is necessary for the proper solution of the problem in hand. If it is M-scaling, the paper should have the columns carefully distinguished and should have room for the answers as they are derived. An extra piece of paper at hand, on which to perform temporary calculations, will keep the record sheet clear and make the results easy to handle. The teacher should do one step at a time and plan not to transfer records from one paper to another unnecessarily. It is difficult to avoid errors when copying.

6. *Neither overvalue nor underestimate the results.* No one process can be a panacea for all educational ills, and Classroom Tests are no exception. They should be like a friend, — someone you know all about and like just the same. The teacher should keep a middle ground, appreciate and value the things that Classroom Tests can do, and use other devices for what they cannot do.

7. *Use Classroom Tests for help and understanding.* Tests should be used to give help to pupils and to provide understanding of pupils for the teacher. The teacher should never forget that he is teaching pupils and not subject matter; that he is testing to teach his pupils in a better way rather than testing merely to test. Tests are not ends in themselves, and never should be. They are one of the many means to an end, and that end is the harmonious development of pupils for worthy living. In so far as Classroom Tests are means to that end they are worth while.

INDEX

- Academic causes of difficulty of school children: mechanical causes, 252-253; lacks-in-knowledge causes, 253; lack of proper attitudes, 254
- Accomplishment Ratio. *See* Achievement Ratio
- Accomplishment Ratio Technic, in Standard testing, 283-286; formula for, 285; essentials of, 286-287; establishment of, for use with Classroom Tests, 287-288; use of Standard Intelligence Tests in, 287-288
- Achievement, need of new point of view toward, 201-202; in relation to ability, 201, 283; in relation to ability of group, 201, 283
- Achievement Ratio, in judging pupils, 255; meaning of, 285-286; finding, from single tests or battery of tests, 288-293; finding, when using composite of several tests, 293-295; interpretations and values of, with Classroom Tests, 295-299
- Achievement Ratio Card, 298
- Administration of tests: True-False Test, 37-47; Judgment Test, 71-76; Selection Tests, 110-117; Association Test, 129-137; Completion Test, 154-157; batteries of tests, 185-186
- Association Test, purposes, 124-126; characteristics, 126; construction, 126-129; administration (by dictation, 129-131 (criticisms of, 131); by blackboard method, 132-133; by mimeograph method, 133-137); sample tests, 135-137, 142-147; scoring, 137-140; disadvantages and values, 140-141
- Base line, determination of, 207-208; placement of, 208; labeling score groups below, 208
- Batteries of tests, advantages, 178-179; disadvantages, 179-180; selection of test types, 180-182; equating length and value of test parts, 182-185; administering, 185-186; scoring, 186-187; samples, 188-197; finding Achievement Ratios from, 288-293
- Blackboard method, for True-False Test, 41-43 (criticisms of, 42-43); for Judgment Test, 73-74; for Selection Tests, 114-115 (criticisms of, 115); for Association Test, 132 (criticisms of, 133); for Completion Test, 154-156 (criticisms of, 156)
- Chance order of elements, in True-False Test, 31-32; in Judgment Test, 70-71; in Selection Tests (Type I, 92-93; Type II, 97; Type III, 104; Type IV, 108); in Association Test, 128; in Completion Test, 153-154
- Characteristics of tests: traditional-type tests, 12-13 (*see also* 166-167); True-False Test, 28; Judgment Test, 67; Selection Tests (Type I, 91; Type II, 95; Type

- III, 102; Type IV, 105-106); Association Test, 126; Completion Test, 148
- Checking by pupils, 48-49, 332-334
- Classification, relation of tests to, 301-303; use of Classroom Tests for, 322-323; within a room (equivalent groups, 323-324; complementary groups, 324-325; special groups, 325); within a grade (equivalent groups, 325-326; complementary groups, 326; special groups, 326); within a school, 327
- Classroom Tests, as supplement to Standard Tests, 22; advantages, 22-24; limitations, 24-25; used for promotion, 303, 320-322; used in conjunction with Standard Tests, 322; used for classification, 322-323 (within a room, 323-325; within a grade, 325-326; within a school, 327); as a teaching device (in general, 329-332; in upper grades, 332-335; in lower grades, 335-338)
- Complementary groups, within a room, 324-325; within a grade, 326
- Completion test, purposes of, 147-148; characteristics of, 148; construction of, 149-154; administration (by dictation, 154; by blackboard method, 154-155 (criticisms, 156); by mimeograph method, 156-157); sample tests, 157, 162-165, 190, 197 (for lower grades, 337); scoring, 158-159; dangers in construction, 159-161
- Composite of M-scale scores, 278-281
- Composite scores, desirability of, 267-269; computation of, 269-281
- Computations for finding percentage of pupils in M-scaling, 273
- Construction of tests: True-False Test, 28-37; Judgment Test, 67-71; Selection tests (Type I, 91-95; Type II, 95-98; Type III, 103-105; Type IV, 106-110); Association Test, 126-129; Completion Test, 149-154; Traditional-Type Test, 167-169; batteries of tests, 180-185
- Correction of tests, by teachers, 47-48; by pupils, 48-49, 332-334
- Dictation method, in True-False Test, 37-40 (criticisms of, 41); in Judgment Test, 71-73; in Selection Tests, 110-113 (advantages and limitations, 113-114); in Association Test, 129-131 (criticisms, 131); in Completion Test, 154
- Difficulties of pupils, physical causes of, 248-250, 262-266; mental causes of, 250-252; academic causes of, 252-254
- Distributions, normal, 214-215; minus-skew, 215-218; plus-skew, 218-221; multimodal, 221-222
- Educational diagnosis, with pupils, 244-245; determination of pupils in need of, 246-248; determination of causes of difficulties of pupils, 248-255
- Equivalent groups, within a room, 323-324; within a grade, 326
- Failure, in relation to Achievement Ratios, 255; meaning of, 308, 320-321
- Fixed standard, 202-204; construction of scale with, 204-205
- Formulas, for True-False Test, 50-52; for finding test range, 207; for finding percentages for M scores, 272; for finding composite,

- 281; for finding Achievement Ratios with Standard Tests, 285; for finding Achievement Ratios with Classroom Tests, 288
- Frequency surface, construction of, 205-211; table for finding divisors in constructing, 206; uses, 210-211, 255; illustration of completed, 211, 248, 268
- Graph of question difficulty, 241, 242, 249
- Graphic presentation of results, 334-335
- Group intelligence tests, described, 18-20; use of, in Achievement Ratio Technic for Classroom Tests, 288-290; precautions in selecting, 290; precautions in giving, 291; precautions in scoring, 291; precautions in special retests, 291-292; given to new pupils, 292-293
- Health disorders affecting school children, 250, 262-266
- Improvement of teaching, 245-246; question analysis for, 255-256, 258; in teaching-skills, 256-257; in teaching-knowledge, 257-259; in teaching-attitudes, 259-261
- Intelligence Quotients converted to Mental Age for M-scaling, 292
- Intelligence Tests, history of development, 18-20; use in ratio technic, 20, 287-288; for determining mental ability, 251; use in ratio technic with Classroom Tests, 288; use to reinforce teacher's judgment, 307
- Interest aroused by tests, 254
- Interpretation of curves: normal distribution, 214-215; minus-skew, 215, 216, 217, 218; plus-skew, 218-219, 219-220, 220, 221; multimodal, 222
- Judging pupils by Achievement Ratios, 295-299
- Judgment Test, characteristics of, 67; construction of, 67-71; administration of (by dictation, 71-73; by blackboard method, 73-74; by mimeograph method, 74-76); samples of, 75, 84-88, 188, 195; scoring, 76-82; values of, 83
- Lack of proper attitudes as causes of difficulties in pupils, 254
- Lacks-in-knowledge as causes of academic difficulties of pupils, 253
- Letter grades, transmuted from scores, 303-313; typical systems of, 304; meaning of divisions, 306-307
- Marking, 201-202, 283, 301-303
- Measurement, in education, 3; of memory, 4; of teacher's efficiency, 4-5
- Mechanical causes of academic difficulties of pupils, 252-253
- Mental Ages, use for Standard M scores, 290; found from Intelligence quotients, 292
- Mental causes of difficulties of pupils, 250-252
- Mental hygiene, applications of, in school, 264-266
- Mimeograph method, for True-False Test, 43-47 (samples of, 45, 46, 57-65, 189, 191-194); for Judgment Test, 74-76 (samples of, 75, 76, 84-88, 188); for Selection Tests, 115-117 (samples of, 93, 98, 104-105, 108, 119-123, 336-337); for Association Test, 133-137 (samples of,

- 135-137, 142-146); for Completion Test, 156-157 (samples of, 157, 162-165)
- Minus-skew curves, 215-218
- M scale, distinguished from T scale, 269-270; finding values for percentages, 275; table of values, 276; figure showing derivation of, 278; revised M-scaling when using composite M-scores as basis, 294-295; used as rating system, 303
- M-scores, formula for finding, 272; finding values for percentages, 275; tables of M-scale values for, 276; relation to promotion, rating, and classification, 279; interpretation of, 281; use in ratio technic, 287-288; derived from Standard Intelligence Tests, 288-293
- Multimodal curves, 221-222
- Normal distribution, 214-215; interpretation of, 215
- Norms, 19-20
- Percentages, quick way of finding, 273-275
- Percentile grades, inaccuracies in, 201-202; transmutation from scores, 313-322
- Percentile groupings for letter-grade systems, 307-309
- Physical causes of difficulties of pupils, 248-250, 262-266
- Plus-skew curves, 218-221
- Promotion, relation of tests to, 301-303; contrasted with failure, 308, 320-321; use of tests for, 320-322
- Pupil correction of tests, 48-49, 332-334
- Purposes of tests: True-False Test, 28; Judgment Test, 66-67; Selection Tests, 89-90 (Type II, 95; Type III, 102; Type IV, 105-106); Association Test, 124-126; Completion Test, 147-148; batteries of tests, 178-179
- Question analysis, importance of, 226-227; for diagnostic purposes, 247-248; for improvement of teaching, 255-256, 258
- Question difficulty, determination for True-False Tests, 227-231; preparation of tally sheet, 227; tallying, 227-228; sample tally sheet, 228; rearrangement in graphical form, 228-231; determination for variably scored tests, 231-240; preparation of tally sheet, 231-232; tallying, 233-234; checking, 235-236; sample sheet, 236; rearrangement in graphical form, 237-240; determination for batteries of tests, 240-243
- Question-difficulty graph, 230, 241, 242, 249
- Rank Order, samples of, 271; placing scores in, 270-271, 310-311; of scores in letter-grade rating system, 310
- Rating, relation of tests to, 301-303
- Rating of pupils, according to individual ability, 201-202, 283; according to ability of group, 201-202, 283
- Raw Score, range for various types in batteries, 184-185; distribution of, 201-212; sample distribution of, 211; used in determining question difficulty, 226-241; used in M-scaling, 269-281; needed for identification of pupils, 278; transmuted to letter-grades, 303-313; transmuted


- to percentile grades, 313-320;
space for, on test papers, *see*
Mimeograph Method, samples of
- Sample computations for finding
M scores, 272-273, 275-278
- Sample tests: True-False, 45-46,
57-65, 189, 193-194, 196; Judg-
ment, 75, 84-88, 188, 195;
Selection Tests (Type I, 93, 119-
121, 123, 197; Type II, 98;
Type III, 104-105; Type IV,
108, 121-123, 194-195); Associa-
tion, 135-137, 142-146; Com-
pletion, 157, 162-165, 190, 197;
batteries of tests, 188-197; for
lower grades, 336-338
- Samples of rank-order distribu-
tions, 271
- School marks, 201-204, 301-310
- Score range, determination for mak-
ing frequency surface, 206-207
(formula for, 207); determina-
tion for percentage groupings,
311-312
- Scores, tabulation of, 52-53; trans-
mutation to letter grades, 303-
313; transmutation to percen-
tile ratings, 313-322
- Scoring: by teachers, 47-48; True-
False Test, 47-53; by pupils, 48-
49, 332-334; Judgment Test,
76-82; Selection Tests, 117-118
(Type I, 93-95; Type II, 99-
102; Type III, 105; Type IV,
117-118); Association Test, 137-
140; Completion Test, 158-159;
Traditional-Test types, 169-175;
batteries of tests, 186-187
- Scoring key, for True-False Tests
(preparation of, 47; use of, 47-
48; sample of, 48); for Judg-
ment Test (construction, 76-78;
sample, 79-80); sample of, for
battery of tests, 191-193
- Selection of subject matter, for
True-False Test, 28; for Judg-
ment Test, 68; for Selection
Tests (Type I, 91; Type II,
95-96; Type III, 103; Type IV,
106); for Association Test, 127;
for Completion Test, 149; for
Traditional-Test types, 167-168
- Selection Test, uses of, 89-90;
varieties, 90; Type I, two-col-
umn selection (construction,
91-95; special considerations,
93-95; sample of, 93, 197);
Type II, rearrangement by se-
lection (construction, 95-98;
sample of, 98; scoring directions,
99-102); Type III, regrouping
by selection (construction, 102-
104; special scoring directions,
105; sample of, 104-105); Type
IV, selection of related from
unrelated facts (construction,
105-108; sample of, 108, 194, 195);
administration (by dictation,
110-113 (advantages and limita-
tions, 113-114); by blackboard
method, 114-115 (criticisms,
115); by mimeograph method,
115-117); scoring, 117-118; sam-
ple for lower grades, 337, 338
- Self-analysis, 245
- Self-improvement, 245
- Self-rating record cards, 298, 335
- Signs of health disorders and
physical defects of school chil-
dren, 262-264
- Smoothed curve, 213-214; example
of, 214
- Special groups, within a room,
325; within a grade, 326
- Standard Deviation, 277-278
- Standard M scores, 288-291; use
of Mental Age for finding, 290
- Standard Tests, development of,
16-20; standardization of, 18-

- 20; disadvantages of, 20-22; use for promotion, 303; used in conjunction with Classroom Tests for promotion, 322
- Standards, fixed, 202-204; varying, 202-203
- Tabulation of test scores, 52-53, 205-211
- Tally sheet of question difficulty, for True-False Tests (preparation of, 227; sample, 228); for variably scored tests (preparation, 231, 232; sample, 236)
- Teacher correction, 47-48
- Teaching-attitudes, improvement of, 259-261
- Teaching-knowledge, improvement of, 257-259
- Teaching-skills, improvement of, 256-257
- Test range, determination of, 206-207, 311-314
- Tests, use of, for examinations, 5; for review and recall, 5-6; for placement and classification, 6-7, 322-323; for classification (within a room, 323-325; within a grade, 325-326; within a school, 327); for diagnosis, 7; for comparison, 7-8; to increase worth of learning, 8-9; to give objective standards, 9; to improve teaching, 9-10, 245-246, 255-261; combining results of several, 267-281; for rating, 283; relation of, to rating, promotion, and classification, 301-303; for promotion, 320-322
- Traditional school test, range of use, 12; advantages, 13; disadvantages, 13-15, 166-167; making effective (selecting objectives and range of subject matter, 167-168; judging length of test and value of elements, 168-169; setting standards for scoring, 169-170; assignment of values to standards, 170-173); scoring, 173-175; uses and limitations, 175-176; use of percentile scale in, 201-230
- Transmutation of scores, to letter grades (placing scores in rank order, 310-311; determining score range for percentage groups, 311-312; application of grades to scores, 312-313); to percentile ratings (determination of comparable score ranges for the letter groups, 313-314; determination of range differences, 314; division of percentile differences by raw-score differences, 315; determination of score distance above lowest score, 316; multiplication of unit distances by proportion found, 316; addition to lowest point of percentile range, 317; successive additions to lowest percentile score, 318; special considerations of E-score groupings, 319-320; completed transmuted raw scores to percentile grades, 320)
- True-False Test, characteristics of, 28; steps in construction of, 28-37; elimination of difficulties, 34-36; administration (by dictation, 37-41; by blackboard method, 41-43; by mimeograph method, 43-47); samples of, 45-46, 57-65, 189, 193-194, 196; scoring, 47-53; defense of, 53-54; use of, 54-56
- T scale, distinguished from M scale, 269-270; description, 277-278
- Varying standards, 202-203

057 0140

3 5111 00057 2440

Date Due



Je 24 '37

1937

237

N 21 '39

163

N 20 40

NO 30'62

36

Feb 8 '61

~~JYL 1962~~

~~AE 26'63~~

~~JY 8'64~~

JE 9 '65

SE 100

WITHDRAWN FROM
OHIO NORTHERN UNIVERSITY
LIBRARY

371.26

R96

UCLC

Heterick Memorial Library
Ohio Northern University

Ada, Ohio 45810



08-BZL-400

